AD NUMBER

ADB176689

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies only; Proprietary Info.; Specific Authority; Jul 93. Other requests shall be referred to RL [OCTM] , Griffiss AFB, NY 13441-4514.

AUTHORITY

AFRL/IFOIP ltr, 15 Jun 2004

AD-B176 689

RL-TR-93-141
Final Technical Report
July 1993

MULTICHANNEL SYSTEM IDENTIFICATION AND DETECTION USING OUTPUT DATA TECHNIQUES

Scientific Studies Corporation

Jaime R. Roman and Dennis W. Davis

93-24068

DTIC
ELECTE
OCT 14 1993
B D

**Rome Laboratory
Air Force Materiel Command
Griffiss Air Force Base, New York**
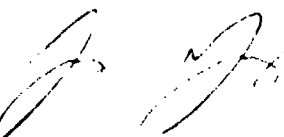
RL-TR-93-141 has been reviewed and is approved for publication.

APPROVED: *[signature]*

JAMES H. MICHELS
Project Engineer

FOR THE COMMANDER: *[signature]*

JAMES W. YOUNGBERG, Lt Col, USAF
Deputy Director
Surveillance and Photonics Directorate

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL (OCTM) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1993	Final Jul 92 – Jan 93

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
MULTICHANNEL SYSTEM IDENTIFICATION AND DETECTION USING OUTPUT DATA TECHNIQUES	C - F30602-92-C-0099 PE - 65502F PR - 3005 TA - RC WU - 63
6. AUTHOR(S) Jaime R. Roman, Dennis W. Davis	

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Scientific Studies Corporation 2250 Quail Ridge Palm Beach Gardens FL 33418	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Rome Laboratory/OCTM 26 Electronic Parkway Griffiss AFB NY 13441-4514	10. SPONSORING/MONITORING AGENCY REPORT NUMBER RL-TR-93-141

11. SUPPLEMENTARY NOTES

Rome Laboratory Project Engineer: Jim Michels/OCTM/(315)330-4431.

12a. DISTRIBUTION/AVAILABILITY STATEMENT USGO agencies only; specific authority: ~~DOD FAR SUP 252.227-7013, ALT II~~. Other requests RL/OCTM, 26 Electronic Parkway, Griffiss AFB NY 13441-4514. *July 93*	12b. DISTRIBUTION CODE PROPRIETARY INFORMATION

13. ABSTRACT (Maximum 200 words)

In multichannel identification and detection (or model-based multichannel detection) problems the parameters of a model are identified from the observed channel process, and the identified model is used to facilitate the detection of a desired signal in the observed process. A model-based multichannel detection algorithm was developed in the context of an innovations-based detection algorithm (IBDA) formulation for surveillance radar system applications. The state space model class was adopted to model the vector channel process because it is more general than the time series model class used in most analyses to date. An IBDA methodology was developed based on an algorithm which uses output data directly and offers computational and performance advantages over alternative techniques. A computer simulation was developed to validate the methodology and the algorithm, and to carry out performance assessments. Simulation results indicate that the algorithm is capable of discriminating between the null hypothesis (clutter plus noise) and the alternative hypothesis (signal plus clutter plus noise). In summary, the applicability of the approach to radar system problems has been established.

14. SUBJECT TERMS Multichannel systems, Identification, Detection			15. NUMBER OF PAGES 132
			16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT SAR

TABLE OF CONTENTS

IV

LIST OF FIGURES

1.0 INTRODUCTION

In multichannel identification problems the outputs of multiple channels (or sensors) are available, and it is desired to identify the parameters of an analytical model to represent the phenomena being observed via the channel outputs. Similarly, in multichannel detection problems the outputs of multiple channels are available, and it is desired to determine the presence (or absence) of a desired signal component in the channel data. In the combined problem of multichannel identification and detection a model is estimated for the phenomena being observed via the channel outputs, and the identified model is used to facilitate the detection of a desired signal in the channel output data. Multichannel identification and detection is thus referred to also as model-based multichannel detection. In all of these problems the channel data is available simultaneously over many channels of the same type, or over many distinct channels (each channel corresponding to a different sensor type).

This report is a summary of the work carried out in Phase I of this program. Specifically, the development of state space algorithms for model-based multichannel detection in the context of surveillance radar system applications is addressed. In surveillance radar systems (radar arrays) the channels correspond to separate antenna apertures (or elements of a single aperture array). The desired signal may or may not be present in the channel output data at any given time. The data in each channel generally includes noise (broadband interference) as well as "clutter" (narrowband interference), with low signal-to-clutter ratio and, possibly, low signal-to-noise ratio also. Model-based detection methods must discriminate between the condition of target embedded in clutter and noise, and the condition of clutter and noise only.

1

Figure 1-1 illustrates a radar array system consisting of multiple subarrays or array elements. The output of each subarray (or each individual array element) is a complex-valued, scalar, digital sequence, denoted as $\{x_i(n)\}$. The collection of the J scalar sequences is arranged into a J-dimensional vector, $\{\underline{x}(n)\}$, which is input to a multichannel processor (not shown in the figure).

Channel No. 1

Channel No. J
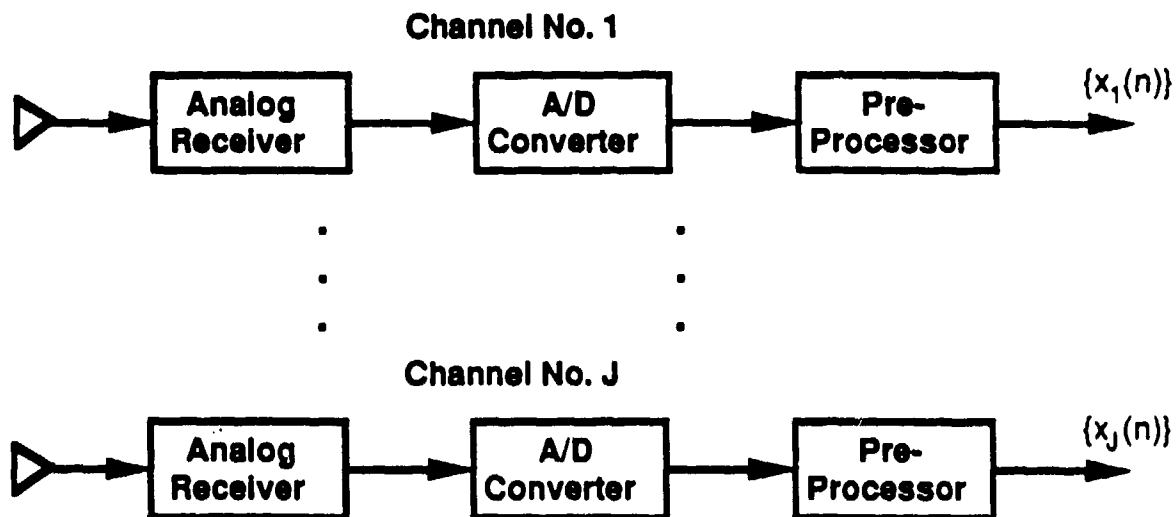
Figure 1-1. Radar array with J subarrays or individual elements.

In Phase I the multivariate (multiple input, multiple output) state space model class was adopted to represent the multichannel radar data, and new system identification techniques were applied to estimate the model parameters. The modeling of the complex-valued pre-processed radar signals for multichannel detection using the state space model class is one of the contributions of this work. State space models have been used in the context of target tracking (where the detected radar signal is processed further to estimate a trajectory) and for the determination of weights in antenna array sidelobe canceling and related problems, but not for multichannel detection. Model-based detection has been carried out using the more-restricted time series models

(Michels, 1991; Metford and Haykin, 1985), which are included within the class of state space models and can be represented as such.

The state space identification algorithm on which the methodology developed in this program is based has several unique features. Foremost among these, the algorithm operates on output data directly to generate estimates of the parameters of a state space model (without computing output correlation matrices). This feature of the algorithm results in reduced dynamic range requirements in comparison with state space algorithms that operate on correlation matrices. The algorithm belongs to the class referred to as subspace methods because the fundamental operation of the algorithm is to decompose the vector space spanned by the channel output data into signal and noise subspaces. Implementation of this fundamental operation is carried out using the QR decomposition and the singular value decomposition (SVD), which are stable numerical techniques. This identification algorithm is new; it is scheduled to appear in the open literature late this year (Van Overschee and De Moor, 1993).

An important distinction in the context of radar system applications is that the vector random processes which represent the channel data are complex-valued processes in most cases. Most time series techniques and models have been formulated for complex as well as real processes. The same, however, cannot be said about state-space techniques; state-space methods and results available in the literature have been defined almost exclusively for the case of real-valued processes, including the algorithm of Van Overschee and De Moor (1993). In Phase I the Van Overschee-De Moor algorithm was extended to the case of complex-valued processes, which is the formulation presented in this report.

3

A computer simulation was generated as part of this program to validate the methodology and the algorithms, and to carry out simulation-based analyses. This software was exercised with simulated multichannel data generated at RL, and the modeling and identification results compare favorably with the results obtained at RL using auto-regressive models.

In summary, the analytical and simulation results obtained in this program indicate that the SSC algorithm and methodology for model-based multichannel detection has the potential to result in significant advances for radar system applications.

1.1 Notation

Vector variables are denoted by underscored lower-case letters (including Greek letters). Matrices are denoted by upper-case letters (including Greek letters). Some scalars (such as the order of the state variable model) are denoted also by upper-case letters. Vector spaces are denoted by upper-case script letters, such as $\mathcal{V}$. The expectation operator is denoted as $E[\cdot]$; superscript T and H are used to denote the matrix and vector transpose and the Hermitian transpose operators, respectively; and an asterisk (*) denotes the complex conjugate operator. I_M denotes an M-dimensional identity matrix, $O_{N,J}$ denotes an $N \times J$ null (zero) matrix, O_M denotes an M-dimensional (square) null matrix, and $\underline{0}_M$ denotes an M-dimensional zero vector. $|A|$ denotes the determinant of matrix A; A^{-1} denotes the inverse of matrix A; A^t denotes the pseudoinverse of A; range(A) denotes the range (column space) of A; rank(A) denotes the rank of A; $A(i,j)$ and a_{ij} are both used to denote the (i,j)th element of matrix A; and dim($\mathcal{V}$) denotes the dimension of vector space $\mathcal{V}$. A caret ($^\wedge$) over a variable denotes an estimate of the variable, a bar (—) over a variable is used to represent the mean of the variable, and ln(a) denotes the natural

4

logarithm of a. The symbol $\perp$ denotes "is orthogonal to;" $\cap$ denotes intersection of two vector spaces; $\oplus$ denotes the direct sum of vector spaces; $\forall$ denotes "for all;" and $\in$ denotes "is an element of."

Where possible, the symbols used herein to represent variables match the symbols used by Michels (1991) to facilitate enhancing the software available at Rome Laboratory (RL) with the techniques developed in this program. This philosophy forces the use of non-standard symbols to represent the parameters of a state variable model. Of course, notational convention should not be a major issue provided all symbols are defined appropriately. However, it is important to mention this point in order to avoid possible confusion on the part of the reader.

1.2 Report Overview

An introduction to the model-based multichannel detection problem is presented in Section 2.0. This section includes also the definition of the state space model class and several related concepts, including the backward model associated with a forward model, and the innovations representation for a random process. The parameter identification algorithm is presented in Section 3.0, and the algorithm proof provided differs significantly from the proof given by Van Overschee and De Moor (1993). In fact, the algorithm proof given here is simpler and easier to follow. As mentioned earlier, this algorithm is the backbone of the Scientific Studies Corporation (SSC) multichannel detection approach. Kalman filtering of the channel data to generate the innovations sequence is discussed in Section 4.0. The innovations sequence is fed to a likelihood ratio detector which generates the detection decision, as described in Section 5.0. A discussion of the software generated in the program is presented in Section 6.0,

along with several simulation results. Section 7.0 includes the main conclusions and recommendations borne out of this Phase I. Appendix A presents a methodology for generating the state space representation of three conventional time series models (moving-average, auto-regressive, and auto-regressive moving-average). Appendix B presents the quotient singular value decomposition (QSVD) for matrix pairs, as required in Section 3.0.

2.0 MODEL-BASED MULTICHANNEL DETECTION

The model-based approach to multichannel detection involves processing the channel data to identify the parameters of a model for the multichannel system, and determination of a detection decision utilizing the identified parameters to filter the channel data. Model parameters can be identified on-line, as the channel data is received and processed. Alternatively, the model parameters can be identified off-line for various conditions and stored in the processor memory to be accessed in real-time as required.

There are two general classes of linear parametric models for vector random processes: time series models and state space models. Time series models include moving-average (MA) models, auto-regressive (AR) models, and auto-regressive moving-average (ARMA) models. State space models are more general than time series models; in fact, MA, AR, and ARMA models can be represented by state space models (Appendix A). In the state space literature, the determination of the model parameters based on output data (and, sometimes, input data also) is referred to as a stochastic identification or a stochastic realization problem.

Time series models have been applied to the multichannel detection problem, and the performance results obtained provide encouragement for further research (see, for example, Michels, 1991, and the references therein). The results obtained by Michels (1991) assume that the multichannel output process can be modeled as a vector AR process. Given the generality of state-space models and the wealth of results available in the state-space literature, the state space model class was selected in Phase I to represent the multichannel signals in the model-based multichannel detection problem for radar systems.

In the case of time series models, two types of model parameter estimation algorithms have been established in the literature: (a) algorithms which operate on channel output correlation matrices, such as the extended Levinson algorithm (Anderson and Moore, 1979), and (b) algorithms which operate on the channel output data directly (without the need to compute channel output correlation matrices), such as the Levinson-Wiggins-Robinson algorithm (Wiggins and Robinson, 1965) and the Strand-Nuttall algorithm (Strand, 1977; Nuttall, 1976).

In the case of state-space models, most of the existing algorithms operate on channel output correlation matrices, such as the stochastic realization approach developed by Akaike (1974, 1975). This limitation is due, in a large part, to the fact that the structure of state space models is more general than the structure of time series models, and the increase in generality has presented a significant challenge to the development of algorithms that operate on channel output data directly. Very recently, however, Van Overschee and De Moor (1993) have defined a state space stochastic realization algorithm which avoids the computation of channel output correlation matrices. Furthermore, this algorithm can be implemented using robust numerical techniques. The Van Overschee-De Moor algorithm was adopted in Phase I to solve the parameter identification problem.

2.1 Multichannel Detection

Detection problems in the context of radar systems can be postulated as hypothesis testing problems, where a choice has to be made among two or more hypotheses. The detection problems addressed in this report involve the following two hypotheses:

8

H_0: Desired signal is absent

H_1: Desired signal is present

H_0 is referred to as the null hypothesis, and H_0 is the alternative hypothesis. The model-based approach to the multichannel detection problem is couched on the assumption that the vector random process at the output of the channels can be represented as the output of a linear system (filter) under each of the two hypotheses, and that a unique parametric model corresponds to each hypothesis. Furthermore, the two parametric models (one for each of the two hypotheses) must be sufficiently different to allow selection of the correct hypothesis by the evaluation of measures that are sensitive to those differences.

A particular measure that has produced robust experimental results in the model-based detection context (Metford and Haykin, 1985) is the log-likelihood ratio (LLR) test. This test is the result of solving the hypothesis testing problem using the Neyman-Pearson criterion. The LLR test in the context of model-based detection is calculated using the innovations sequence at the output of each of the two linear filters. This presents practical and implementation advantages.

Figure 2-1 illustrates the architecture of an on-line innovations-based multichannel detector. In the case of a radar array system, each of **J** radar receiver channels collects the electromagnetic energy arriving at its aperture, and processes it to generate a discrete-time random sequence, denoted as $\{x_i(n)\}$, which contains the desired information. The **J** random sequences $\{x_i(n)\}$ are represented in vector form as $\{\underline{x}(n)\}$. Michels (1991) has formulated the binary detection problem for multichannel systems. Specifically, the null hypothesis, H_0, corresponds to the case of

9

clutter and noise present in the observation process $\{\underline{x}(n)\}$, and the alternative hypothesis, $\mathbf{H_1}$, corresponds to the case of signal, clutter, and noise present in the observation process $\{\underline{x}(n)\}$. That is, the detection decision must be made between the following two models,

(2-1a) $\mathbf{H_0}$: $\underline{x}(n) = \underline{c}(n) + \underline{w}(n)$ $n \geq n_0$

(2-1b) $\mathbf{H_1}$: $\underline{x}(n) = \underline{s}(n) + \underline{c}(n) + \underline{w}(n)$ $n \geq n_0$

where n_0 denotes the initial observation time, $\{\underline{c}(n)\}$ denotes the clutter process, $\{\underline{w}(n)\}$ denotes all the array channel noise processes, and $\{\underline{s}(n)\}$ denotes the desired signal (target) process. In the model-based approach pursued herein, a distinct state variable model is associated with each of the two hypotheses, and a Kalman filter is designed for each model. Each Kalman filter processes the observation sequence $\{\underline{x}(n)\}$ to generate a vector innovations sequence: $\{\underline{\varepsilon}(n \mid \mathbf{H_0})\}$ denotes the innovations sequence at the output of the null hypothesis filter, and $\{\underline{\varepsilon}(n \mid \mathbf{H_1})\}$ denotes the innovations sequence at the output of the alternative hypothesis filter. These innovations sequences are used in a likelihood ratio test with a pre-stored threshold to carry out the detection decision.

As indicated in the detection configuration of Figure 2-1, the two filters can be determined in real-time by processing the observation sequence for a prescribed time interval. This approach provides the most adaptability, but may present a large computational burden for some applications. It also presents conceptual challenges, such as real-time determination of model order for each of the two filters. Alternatively, the filter design can be carried out off-line for each of the two hypotheses, and the resulting filter design implemented in the real-time

10

configuration. This alternative approach is less robust to changes in the operational environment, but requires a simpler processor architecture, which is important in many real-time applications. Careful design of the filters off-line using adequate simulated and real data can lead to acceptable performance. Also, many pairs of fixed filters may be designed to cover distinct operational conditions. The filter for the alternative hypothesis will be of higher order than the filter for the null hypothesis because the observation process for the alternative hypothesis has more information (the signal component).
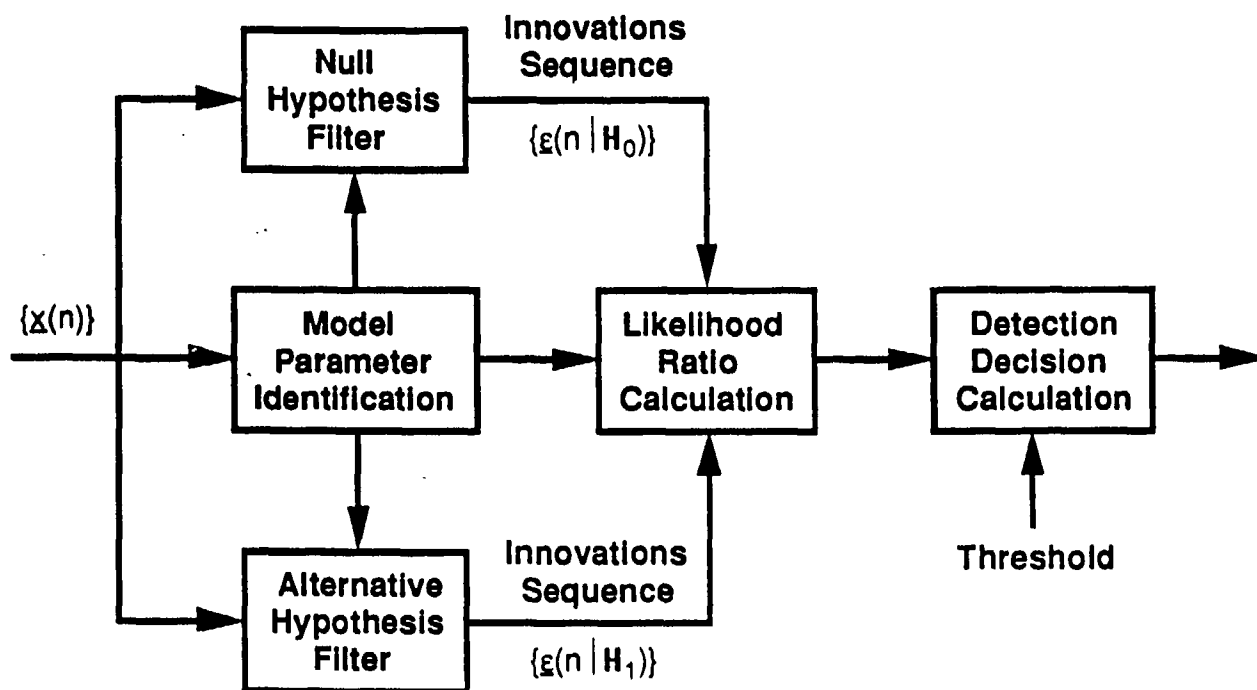
Figure 2-1.   Innovations-based multichannel detector with on-line parameter identification.

Michels (1991) has developed a likelihood ratio calculation and detection decision model which are compatible with the formulation adopted herein. Both of these capabilities are

available at RL, and, where appropriate, the methodology presented in this report is compatible with these capabilities.

2.2 State Space Model

The class of multiple-input, multiple-output state variable models can represent effectively the channel output process for radar applications. Consider a discrete-time, stationary, complex-valued, zero-mean, Gaussian random process $\{\underline{x}(n)\}$ defined as the output of the following state space model representation for the system giving rise to the observed process:

(2-2a) $\qquad \underline{y}(n+1) = F\underline{y}(n) + G\underline{u}(n) \qquad\qquad\qquad n \geq n_0$

(2-2b) $\qquad \underline{x}(n) = H^H\underline{y}(n) + D^H\underline{w}(n) \qquad\qquad\quad n \geq n_0$

(2-2c) $\qquad E[\underline{y}(n_0)] = \underline{0}_N$

(2-2d) $\qquad E[\underline{y}(n_0)\underline{y}^H(n_0)] = P_0$

Here $n = n_0$ denotes the initial time (which can be adopted as 0 since the system is stationary). Also, $\underline{y}(n)$ is the N-dimensional state of the system with $\underline{y}(n_0)$ a Gaussian random vector; $\underline{u}(n)$ is the J-dimensional, zero-mean, stationary, Gaussian, white input noise process; and $\underline{w}(n)$ is the J-dimensional, zero-mean, stationary, Gaussian, white measurement noise process. The output (or measurement) process $\{\underline{x}(n)\}$ is also a J-dimensional vector process. Matrix F is the $N \times N$ system matrix, G is $N \times J$ input noise distribution matrix, H^H is the $J \times N$ output distribution matrix, D^H is the $J \times J$ output noise distribution matrix, and P_0 is the correlation matrix of the initial state. All these matrices are time-invariant. Matrix P_0 is Hermitian and positive definite.

12

System (2-2) is assumed to be asymptotically stable, which means that all the eigenvalues of matrix F are inside the unit circle. Also, system (2-2) is assumed to be reachable and observable, which implies that the dimension N of the state vector (also the order of the system) is minimal (Anderson and Moore, 1979). That is, there is no system of lesser order which has identical input/output behaviour. The output distribution matrices are defined with the conjugate operator in order to have notation consistent with that of the single-output system case, where both H and D become vectors, and nominally vectors are defined as column vectors.

The input noise process correlation matrix is given as (all matrices defined hereafter have appropriate dimensions)

(2-3a) $E[\underline{u}(k)\underline{u}^H(k)] = R_{uu}(0) = Q$ $k \geq n_0$

(2-3b) $E[\underline{u}(k)\underline{u}^H(k-n)] = R_{uu}(n) = [0]$ $k \geq n_0$ and $n \neq 0$

and the output noise process correlation matrix is given as

(2-4a) $E[\underline{w}(k)\underline{w}^H(k)] = R_{ww}(0) = C$ $k \geq n_0$

(2-4b) $E[\underline{w}(k)\underline{w}^H(k-n)] = R_{ww}(n) = [0]$ $k \geq n_0$ and $n \neq 0$

Notice that matrices Q and C are Hermitian (that is, $Q^H = Q$, and $C^H = C$). Matrix Q is at least a positive semidefinite matrix since it is an auto-correlation matrix (all the eigenvalues of a positive semidefinite matrix are non-negative), and matrix C is assumed to be positive definite (this can be relaxed to positive semi-definite, but positive definiteness is more realistic since in the radar problem $\underline{w}(n)$ represents channel noise and other such noise processes which are independent from channel to channel).

13

In the most general form for this model the input and output noise processes are correlated, with a cross-correlation matrix defined as

$$(2\text{-}5a) \qquad E[\underline{u}(k)\underline{w}^H(k)] = R_{uw}(0) = S \qquad\qquad k \geq n_o$$

$$(2\text{-}5b) \qquad E[\underline{u}(k)\underline{w}^H(k\text{-}n)] = R_{uw}(n) = [0] \qquad\qquad k \geq n_o \text{ and } n \neq 0$$

In general, matrix S is not Hermitian. Both the input and output noise processes are uncorrelated with the present and past values of the state process, and this is expressed in terms of cross-correlation matrices as

$$(2\text{-}6a) \qquad E[\underline{y}(k)\underline{u}^H(k\text{-}n)] = R_{yu}(n) = [0] \qquad\qquad k \geq n_o \text{ and } n \geq 0$$

$$(2\text{-}6b) \qquad E[\underline{y}(k)\underline{w}^H(k\text{-}n)] = R_{yw}(n) = [0] \qquad\qquad k \geq n_o \text{ and } n \geq 0$$

The correlation matrix of the state is defined as

$$(2\text{-}7) \qquad E[\underline{y}(n)\underline{y}^H(n)] = R_{yy}(n) = P(n) \qquad\qquad k \geq n_o \text{ and } n \geq 0$$

It follows from (2-2a) and the above definitions that the state correlation matrix satisfies the following recurrence relation,

$$(2\text{-}8) \qquad P(n+1) = FP(n)F^H + GQG^H \qquad\qquad n \geq n_o$$

In general, matrix $P(n)$ is Hermitian and positive definite. Since system (2-2) is stationary and asymptotically stable, and since matrix Q is positive definite, then the following steady-state (large n) value exists for the recursion (2-8):

$$(2\text{-}9) \qquad P(n+1) = P(n) = P$$

14

Under steady-state conditions Equation (2-8) becomes a Lyapunov equation for the steady-state correlation matrix, denoted as P:

(2-10) $\dot{P} = FPF^H + GQG^H$

The conditions for steady-state also insure that the solution to Equation (2-10) exists, is unique (for the selected state space basis), and is positive definite (Anderson and Moore, 1979). Matrix P is unique for a given state space basis. However, if the basis of the input noise and/or the basis of the state are changed by a similarity and/or an input transformation, then a different state correlation matrix results from Equation (2-10).

The correlation matrix sequence of the output process $\{\underline{x}(n)\}$ is defined as

(2-11a) $E[\underline{x}(k)\underline{x}^H(k\text{-}n)] = R_{xx}(n) = \Lambda_n$ $\forall\, k$ and $n \geq 0$

(2-11b) $R_{xx}(\text{-}n) = R_{xx}^H(n)$ $\forall\, n$

For a system of the form (2-2), the correlation matrix $R_{xx}(n)$ can be factorized as follows,

(2-12a) $\Lambda_n = R_{xx}(n) = H^H F^{n\text{-}1} \Gamma$ $n > 0$

(2-12b) $\Lambda_n = R_{xx}(n) = \Gamma^H[F^{n\text{-}1}]^H H = \Gamma^H[F^H]^{n\text{-}1} H$ $n < 0$

where $F^{n\text{-}1}$ denotes F raised to the $(n\text{-}1)$th power and Γ denotes the following cross-correlation matrix

(2-13) $\Gamma = E[\underline{y}(n)\underline{x}^H(n\text{-}1)] = R_{yx}(1) = FP(n)H + GSD$ $\forall\, n > 0$

15

The correlation matrix sequence factorization in Equation (2-12) is the key to most correlation-based stochastic realization algorithms. The zero-lag ($n = 0$) output correlation matrix is

(2-14) $R_{xx}(0) = H^H P(n) H + D^H C D = \Lambda_0$

Matrix $R_{xx}(0)$ is Hermitian and at least positive semidefinite. In steady-state, P replaces $P(n)$ in Equations (2-13) and (2-14).

As can be inferred from the above relations, the system parameters $\{F, G, H, D, Q, C, S, P, \Gamma\}$ completely define the second-order statistics (the correlation matrix sequence $\{R_{xx}(n)\}$) of the output process, and it is said that system (2-2) realizes the output correlation matrix sequence. Conversely, the second-order statistics of the output process provide sufficient information to identify the system parameters, although not uniquely. Since the output process has zero mean and is Gaussian-distributed, the second-order statistics define the process completely.

From the system identification (stochastic realization) point of view, the problem addressed herein can be stated as follows: given the output data sequence $\{x(n)\}$ of system (2-2), estimate a set of system parameters $\{F, G, H, D, Q, C, S, P, \Gamma\}$ which generates the same output correlation matrix sequence as system (2-2). Furthermore, the identified parameter set must correspond to a system realization of minimal order (with state vector y of minimal dimension).

It is well known (Anderson and Moore, 1979) that there can exist an infinity of systems (2-2) with the same output correlation matrix sequence. The set of all systems that have the same output correlation matrix sequence is an equivalence class, and any two systems belonging to the set are said to be

16

correlation equivalent (Candy, 1976). For example, the output correlation matrix sequence remains invariant to a similarity transformation applied to the state vector. Similarly, the output correlation matrix sequence remains invariant also to a non-singular transformation applied to the input noise and/or to the output noise. As shown by Candy (1976), the equivalence class of correlation equivalent systems is defined including other operations besides a change of basis.

Based on these comments, the solution to the system identification problem is not unique. It is also true that most of the possible system parameter solutions do not possess desirable properties. There is, however, a solution which has several features of importance. This solution is referred to as the innovations representation for system (2-2), and is discussed in Section 2.3.

In general, the system matrix parameters resulting from the identification algorithm will be represented in a different basis, and should be denoted with a different symbol (say, F_1 instead of F, etc.); nevertheless, the same symbol will be used in this report in order to simplify notation.

Several definitions and notation associated with the input /output behaviour of system (2-2) are important. Consider first the L-term (finite) controllability matrix of system (2-2), C_L; this matrix is defined as an $N \times JL$ partitioned matrix of the form

(2-15) $$C_L = \begin{bmatrix} G & FG & \cdots & F^{L-1}G \end{bmatrix}$$

As is well-known, matrix C_L has rank N (equal to the system order) for $L \geq N$. The controllability matrix maps the input space onto

the state space. Analogously, the L-term observability matrix of system (2-2) is the following $JLxN$ partitioned matrix,

$$(2\text{-}16) \qquad O_L = \begin{bmatrix} H^H \\ H^H F \\ \vdots \\ H^H F^{L-1} \end{bmatrix}$$

and the rank of matrix O_L is equal to N also for $L \geq N$. The observability matrix maps the state space onto the output space. Classical realization theory for the deterministic case is based on the fact that a block Hankel matrix made up of the impulse response matrices (Markov parameters) of a deterministic system can be represented as the product of the observability and controllability matrices. That is,

$$(2\text{-}17) \qquad H_{L,L} = O_L C_L$$

where $H_{L,L}$ is a $JLxJL$ deterministic Hankel matrix with the impulse response matrix $A(i+j-1)$ as its (i,j)th block element (a block Hankel matrix is a matrix in which the (i,j)th block element is a function of $i+j$). This result follows from the definition of the impulse response matrix sequence,

$$(2\text{-}18) \qquad A(n) = H^H F^{n-1} G \qquad\qquad\qquad n \geq 1$$

Notice that the factorization of the impulse response matrix sequence in Equation (2-18) is very similar to the factorization of the correlation matrix sequence in Equation (2-12).

Associated with system (2-2) is a backward time model which is defined from the system model (2-2). Backward time models play

a role in the formulation of a large class of stochastic realization algorithms. The backward time model for system (2-2) is defined as a discrete-time, stationary, complex-valued, zero-mean, Gaussian random process with a state space representation of the form (Faurre, 1976)

$$(2\text{-}19\text{a}) \qquad \underline{s}(n) = F^H \underline{s}(n+1) + \underline{v}_i(n)$$

$$(2\text{-}19\text{b}) \qquad \underline{x}(n) = \Gamma^H \underline{s}(n) + \underline{v}_o(n)$$

where $\underline{s}(n)$ is the N-dimensional state vector, $\underline{v}_i(n)$ is the N-dimensional input noise vector, and $\underline{v}_o(n)$ is the J-dimensional output noise vector. Both noise vectors are uncorrelated in time (white), have mean equal to zero, and are Gaussian-distributed. The L-term observability matrix for the backward system (2-19) is the following JLxN partitioned matrix,

$$(2\text{-}20) \qquad \mathcal{D}_L = \begin{bmatrix} \Gamma^H \\ \Gamma^H F^H \\ \vdots \\ \Gamma^H (F^H)^{L-1} \end{bmatrix}$$

Also of interest is the Hermitian of $\mathcal{D}_L$ with the block columns in reversed order. That is,

$$(2\text{-}21) \qquad \mathcal{B}_L = \overset{\leftrightarrow}{\mathcal{D}}_L^H = \begin{bmatrix} F^{L-1}\Gamma & \cdots & F\Gamma & \Gamma \end{bmatrix}$$

where the dual-point arrow over matrix $\mathcal{D}_L^H$ indicates reversal in the order of the block columns. Notice that matrix $\mathcal{B}_L$ is like a controllability matrix for the matrix pair (F, Γ) in reverse block

column order. Thus, matrix A_c is referred to herein as the L-term reversed dual controllability matrix.

In the context of stochastic realization theory, the significance of the backward model follows from Equation (2-20) and the Hankel matrix of output correlation matrices, as shown next. Define a stochastic Hankel matrix $\mathcal{H}_{L,L}$ as the following JLxJL block matrix,

$$(2\text{-}22) \qquad \mathcal{H}_{L,L} = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \cdots & \Lambda_L \\ \Lambda_2 & \Lambda_3 & \cdots & \Lambda_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_L & \Lambda_{L+1} & \cdots & \Lambda_{2L-1} \end{bmatrix}$$

where the block elements (Λ_i) are the elements of the output correlation matrix sequence, Equation (2-12). It follows from Equations (2-12), (2-16), and (2-22) that

$$(2\text{-}23) \qquad \mathcal{H}_{L,L} = \mathcal{O}_L \mathcal{D}_L^H$$

This equation is fundamental to stochastic realization algorithms, and allows the application of classical deterministic realization algorithms to the stochastic realization problem formulated with output correlation matrices. It also provides insight into the stochastic realization algorithm presented in Section 3.0, even though the algorithm does not require computation of the output correlation matrix sequence.

Other important matrices in stochastic realization theory include the JLxJL "future" and "past" block correlation matrices. These matrices are the correlation matrices of future and past output block vectors defined as

20

$$(2\text{-}24) \qquad \underline{x}_p = \underline{x}(n;n+L\text{-}1) = \begin{bmatrix} \underline{x}(n) \\ \underline{x}(n+1) \\ \vdots \\ \underline{x}(n+L\text{-}1) \end{bmatrix}$$

$$(2\text{-}25) \qquad \underline{x}_F = \underline{x}(n+L;n+2L\text{-}1) = \begin{bmatrix} \underline{x}(n+L) \\ \underline{x}(n+L+1) \\ \vdots \\ \underline{x}(n+2L\text{-}1) \end{bmatrix}$$

With these definitions, the future and past block correlation matrices are given by the following $JL \times JL$ matrices:

$$(2\text{-}26) \qquad \mathcal{R}_{P;L,L} = E[\underline{x}_P \underline{x}_P^H] = \begin{bmatrix} \Lambda_0 & \Lambda_{\text{-}1} & \cdots & \Lambda_{1\text{-}L} \\ \Lambda_1 & \Lambda_0 & \cdots & \Lambda_{2\text{-}L} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{L\text{-}1} & \Lambda_{L\text{-}2} & \cdots & \Lambda_0 \end{bmatrix}$$

$$(2\text{-}27) \qquad \mathcal{R}_{F;L,L} = E[\underline{x}_F \underline{x}_F^H] = \begin{bmatrix} \Lambda_0 & \Lambda_1 & \cdots & \Lambda_{L\text{-}1} \\ \Lambda_{\text{-}1} & \Lambda_0 & \cdots & \Lambda_{L\text{-}2} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{1\text{-}L} & \Lambda_{2\text{-}L} & \cdots & \Lambda_0 \end{bmatrix}$$

where $\mathcal{R}_{F;L,L}$ and $\mathcal{R}_{P;L,L}$ are the future and past block correlation matrices, respectively. Another matrix of interest is the block cross-correlation matrix between the future and the past, which is defined as

21

$$(2\text{-}28) \qquad \mathcal{R}_{P:L,P:L} = E\{\underline{X}_{P:L}\underline{X}_{P:L}^H\} = \begin{bmatrix} \Lambda_L & \Lambda_{L-1} & \cdots & \Lambda_1 \\ \Lambda_{L+1} & \Lambda_L & \cdots & \Lambda_2 \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{2L-1} & \Lambda_{2L-2} & \cdots & \Lambda_L \end{bmatrix} = \tilde{\mathcal{H}}_{L,L} = \mathcal{O}_L \tilde{\mathcal{R}}_L$$

Notice that the block cross-correlation matrix $\mathcal{R}_{P:L,P:L}$ is equal to the stochastic block Hankel with the block columns in reverse order, as indicated in Equation (2-28). For $L \geq N$, equations (2-26)-(2-29) define the correlation structure of system (2-2). In fact, the stochastic realization algorithm of Akaike (1974, 1975) is based on these block correlation matrices.

2.3 Innovations Representation

The innovations representation is a very powerful concept in the theory of linear stochastic systems due to its simplicity and its characteristics. Several texts and papers discuss this concept in detail; in particular, Anderson and Moore (1979) provide a lucid presentation. The discussion herein is adapted mostly from Anderson and Moore (1979).

The innovations representation for a system (2-2) is a discrete-time, stationary, complex-valued, system of the form

$(2\text{-}29a) \qquad \underline{\alpha}(n+1) = F\underline{\alpha}(n) + K\underline{\varepsilon}(n) \qquad\qquad n \geq n_0$

$(2\text{-}29b) \qquad \underline{\chi}(n) = H^H\underline{\alpha}(n) + \underline{\varepsilon}(n) \qquad\qquad n \geq n_0$

$(2\text{-}29c) \qquad \underline{\alpha}(n_0) = \underline{0}_N$

$(2\text{-}29d) \qquad E[\underline{\alpha}(n_0)\underline{\alpha}^H(n_0)] = \Pi(n_0) = \Pi_0 = [0]$

22

(2-29e) $E[\underline{\alpha}(n)\underline{\alpha}^H(n)] = \Pi(n)$ $n > n_0$

(2-29f) $\Pi(n) = \Pi$ as $n \rightarrow \infty$

(2-29g) $R_{\chi\chi}(n) = R_{xx}(n)$ $\forall\, n$

here $\underline{\alpha}(n)$ is the N-dimensional state, $\underline{\chi}(n)$ is the J-dimensional output, and the input process $\{\underline{\varepsilon}(n)\}$ is the innovations process for system (2-2). That is, $\{\underline{\varepsilon}(n)\}$ is a J-dimensional, zero-mean, white Gaussian process with correlation matrix structure given as

(2-30a) $\Omega = E[\underline{\varepsilon}(k)\underline{\varepsilon}^H(k)] = R_{xx}(0) - H^H\Pi H = \Lambda_0 - H^H\Pi H$ $k \geq n_0$

(2-30b) $E[\underline{\varepsilon}(k)\underline{\varepsilon}^H(k\text{-}n)] = [0]$ $k \geq n_0$ and $n \neq 0$

The state correlation matrix $\Pi(n)$ has a steady-state value because the system is asymptotically stable (stationary), and the steady-state value, Π, is obtained as the limiting solution to the following recursion

(2-31a) $\Pi(n+1) = F\Pi(n)F^H + [F\Pi(n)H - \Gamma][\Lambda_0 - H^H\Pi(n)H]^{-1}[F\Pi(n)H - \Gamma]^H$ $n \geq n_0$

(2-31b) $\Pi(n_0) = \Pi_0 = [0]$

Matrix K in Equation (2-29a) is given as

(2-32a) $K = [\Gamma - F\Pi H]\,\Omega^{-1} = [\Gamma - F\Pi H][\Lambda_0 - H^H\Pi H]^{-1}$

(2-32b) $K = GSD\,\Omega^{-1} = GSD[\Lambda_0 - H^H\Pi H]^{-1}$

where the second relation follows from the definitions of Γ in Equation (2-13) and of Ω in Equation (2-30a). In the cases where

23

the inverse of the correlation matrix Ω does not exist, its pseudoinverse is used instead in Equations (2-31) and (2-32).

Matrices F, H, Λ_0, and Γ are as defined for system (2-2). That is, system (2-29) is related to system (2-2). In fact, system (2-29) as defined above is the steady-state innovations representation for system (2-2). This representation has the following important features.

(a) First and foremost, the correlation matrix sequence of $\{\chi(n)\}$ is equal to the correlation matrix sequence of $\{x(n)\}$, as indicated in Equation (2-29g). That is, the processes $\{\chi(n)\}$ and $\{x(n)\}$ are correlation equivalent. This means that the innovations representation is a valid solution to the system identification problem defined herein.

(b) Of all the correlation equivalent representations for a given output correlation sequence, the innovations representation has the smallest state correlation matrix, Π (smallest is meant in the sense of positive definiteness; that is, Π_1 is smaller than Π_2 if $\Pi_2 - \Pi_1$ is a positive definite matrix). This property of the innovations model is significant because the state correlation matrix is a measure of the uncertainty in the state.

(c) The innovations representation is directly related to the steady-state Kalman filter (in the one-step predictor formulation) for system (2-2). In fact, the steady-state Kalman filter for system (2-2) is available immediately upon definition of the steady-state innovations representation, and viceversa.

24

Specifically, matrix K of Equations (2-29a) and (2-31) is the steady-state Kalman gain of the optimal one-step predictor for system (2-2). This is true provided that the eigenvalues of F-KH^H are stable. Thus, the innovations model is defined as above for all processes of the form (2-2), but the steady-state Kalman filter is defined only if F-KH^H is stable.

(d) The process $\{\varepsilon(n)\}$ in Equations (2-29) and (2-30) is correlation equivalent to the innovations sequence of system (2-2), which is the reason for referring to system (2-29) as the "innovations representation" for system (2-2).

(e) The innovations model (2-29) is causally invertible. This means that the present and past of the process $\{\varepsilon(n)\}$ can be constructed from the present and past values of the output process $\{\chi(n)\}$. The converse statement is true also; that is, any causally invertible model is an innovations representation for some system. Causal invertibility of system (2-29) can be demonstrated easily. From Equation (2-29b),

(2-33) $$\varepsilon(n) = -H^H \alpha(n) + \chi(n)$$

Substituting this expression for $\varepsilon(n)$ into Equation (2-29a) results in

(2-34) $$\alpha(n+1) = [F - KH^H]\alpha(n) + K\chi(n)$$

These relations demonstrate the causal invertibility of the innovations model (the input and output variables have traded places).

25

(f) Matrix $F - KH^H$ in the inverted innovations model is a stable matrix. This follows from the fact that the matrix pair (F, H) is observable, and implies that the Kalman filter for system (2-2) is stable also.

(g) The transfer function of the innovations model (2-29) is minimum phase. This is related to the fact that the innovations model is correlation equivalent to system (2-2), and second-order moment information (the output correlation matrix sequence) does not contain any phase information.

(h) The innovations representation for a system of the form (2-2) is unique. Given that the innovations representation has the same output covariance sequence as system (2-2), the fact that it is unique eliminates searching for other representations for system (2-2) with the properties listed herein.

(i) The innovations model (2-29) can be computed from the output correlation matrix sequence of system (2-2). This fact simplifies the parameter identification problem because the set of matrix parameters that must be estimated is reduced to just five: $(F, H, \Gamma, \Pi, \Lambda_0)$ (given these parameter matrices, the innovations covariance, Ω, and the Kalman gain, K, are obtained using Equations (2-30a) and (2-32a), respectively).

All the features listed above are of relevance to the identification approach presented in Section 3.0 because the selected parameter identification algorithm generates the

innovations representation for the given output correlation matrix sequence, following feature (i).

The backward model has an associated backward innovations model which is defined by F, Γ, and the backward Kalman gain. Most of the features (a)-(i) that describe the forward innovations model are valid also for the backward innovations model, with a notable exception of feature (b), which needs to be replaced by the following statement: For each valid correlation equivalent representation for a given output correlation sequence, the state correlation matrix is smaller than the inverse of the state correlation matrix for the backward innovations model. More specifically, let Π_b denote the state correlation matrix for the backward innovations model in steady-state conditions, and let Σ denote the state correlation matrix for any valid correlation equivalent representation of an output correlation sequence. Then, $\Pi_b^{-1} - \Sigma$ is a positive definite matrix. This result provides an upper bound for the state correlation matrix of a correlation equivalent representation, and can be combined with the lower bound established by property (b) of the forward innovations model to give

(2-35) $$\Pi \leq \Sigma \leq \Pi_b^{-1}$$

As before, the inequality between two matrices is intended in the sense of positive semi-definiteness of the matrix difference.

27

3.0 MULTICHANNEL SYSTEM IDENTIFICATION

Identification of the model parameter matrices $\{F, H, \Gamma, \Pi, \Lambda_0\}$ is carried out using the algorithm of Van Overschee and De Moor (1993), extended to the case of complex-valued data. The Van Overschee-De Moor algorithm is based on the predictor space concept of Akaike (1974; 1975), the correlation equivalence results obtained by Faurre (1976), and the balanced stochastic realization approach of Arun and Kung (1990). The algorithm approach is presented herein from a viewpoint which is different from, and simpler than, the presentation given by Van Overschee and De Moor (1993).

3.1 Output Data-Based Algorithm

In comparison with alternative stochastic realization techniques, the Van Overschee-De Moor algorithm adopted herein has several advantages for multichannel detection applications, as listed next.

- Reduced dynamic range with respect to algorithms which require generation of the output correlation matrix sequence (correlation matrices are estimated as sums of products of the data sequence elements, which increases the dynamic range). As such, the algorithm can be viewed as a "square-root" algorithm.

- Identifies the parameters for a model in the state-space class, which is more general than the time series class.

- Belongs to a class of algorithms referred to as "subspace methods." Subspace methods involve the decomposition of the space spanned by the output process

28

into two orthogonal subspaces: one subspace is the space spanned by the "desired component," and the other subspace is spanned by the "noise component." The MUSIC algorithm (Schmidt, 1979; 1981), for example, also belongs to the class of subspace methods.

- An approximately balanced (in the stochastic sense) state space realization is generated, thus providing a built-in and robust mechanism for model order selection.

- Identifies the innovations representation of the system, and generates the Kalman gain directly, without having to solve a nonlinear discrete matrix Riccati equation.

- Approach differs from others in that the states of a Kalman filter for the given sequence are identified first, and then the model parameters are estimated via least-squares.

- Implementation of the algorithm involves the QR decomposition and the quotient SVD (QSVD; also known as the generalized SVD), which are stable numerical methods. Furthermore, the QSVD is applied to matrices of small dimensions.

An algorithm for implementing the QSVD is given in Appendix B for the specific conditions presented in this section.

Consider the channel output sequence $\{\underline{x}(n)\}$. For simplicity, let the initial time $n_0 = 0$. This can be done without loss of generality because the system is stationary. Now define a block Hankel matrix $X_{0,L-1}$ with output sequence vectors assigned as block elements according to the rule $X_{0,L-1}(i,j) = \underline{x}(i+j-2)$; that is,

29

$$(3\text{-}1) \qquad X_{0,L\text{-}1} = \begin{bmatrix} \underline{x}(0) & \underline{x}(1) & \underline{x}(2) & \cdots & \underline{x}(M\text{-}1) \\ \underline{x}(1) & \underline{x}(2) & \underline{x}(3) & \cdots & \underline{x}(M) \\ \underline{x}(2) & \underline{x}(3) & \underline{x}(4) & \cdots & \underline{x}(M\text{+}1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \underline{x}(L\text{-}1) & \underline{x}(L) & \underline{x}(L\text{+}1) & \cdots & \underline{x}(L\text{+}M\text{-}2) \end{bmatrix}$$

Here the first subscript denotes the time index of the first element of the first row, and the second subscript denotes the time index of the first element of the last row. Matrix $X_{0,L\text{-}1}$ has JL rows and M columns, with M >> JL, and JL > N (recall that N is the system order and J is the number of channels). The block row dimension, L, must be selected so that $L \geq N + 1$. Similarly, define another JLxM block Hankel matrix $X_{L,2L\text{-}1}$ with output sequence vectors assigned as block elements according to the rule $X_{L,2L\text{-}1}(i,j) = \underline{x}(i+j-2+L)$; that is,

$$(3\text{-}2) \qquad X_{L,2L\text{-}1} = \begin{bmatrix} \underline{x}(L) & \underline{x}(L\text{+}1) & \underline{x}(L\text{+}2) & \cdots & \underline{x}(L\text{+}M\text{-}1) \\ \underline{x}(L\text{+}1) & \underline{x}(L\text{+}2) & \underline{x}(L\text{+}3) & \cdots & \underline{x}(L\text{+}M) \\ \underline{x}(L\text{+}2) & \underline{x}(L\text{+}3) & \underline{x}(L\text{+}4) & \cdots & \underline{x}(L\text{+}M\text{+}1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \underline{x}(2L\text{-}1) & \underline{x}(2L) & \underline{x}(2L\text{+}1) & \cdots & \underline{x}(2L\text{+}M\text{-}2) \end{bmatrix}$$

Matrices $X_{0,L\text{-}1}$ and $X_{L,2L\text{-}1}$ represent the "past" and the "future", respectively, of the output process. Also let $\mathcal{X}^-$ denote the vector space spanned by the past of the process $\{\underline{x}(n)\}$, and $\mathcal{X}^+$ denote the vector space spanned by the future of the process.

The algorithm is based on the decomposition of the process future (as represented by matrix $X_{L,2L-1}$) into two orthogonal subspaces (herein orthogonality is invoked under the standard unitary inner product for complex vector spaces, with identity metric). In such a decomposition, one subspace is the space spanned by the process past, X^- (as represented by matrix $X_{0,L-1}$), and the second subspace is the space spanned by the noise process. Let $\mathcal{W}$ denote the space spanned by the noise process $\{\underline{w}(n)\}$. Then, the desired decomposition of X^+ is as follows:

$$(3-3) \qquad X^+ = X^- \oplus \mathcal{W}$$

where $\oplus$ denotes the direct sum, and $X^- \perp \mathcal{W}$ (since the present state and the present measurement noise are uncorrelated). In matrix notation, the desired decomposition of $X_{L,2L-1}$ is expressed as

$$(3-4) \qquad X_{L,2L-1} = X_{F \| P} + X_{F \perp P}$$

where the JLxM matrix $X_{F \| P}$ is the projection of the row space of $X_{L,2L-1}$ (the future) onto the row space of $X_{0,L-1}$ (the past), and the JLxM matrix $X_{F \perp P}$ is the projection of the row space of $X_{L,2L-1}$ onto the complement of the row space of $X_{0,L-1}$. Akaike (1974; 1975) has demonstrated that since the order of the state space model is N, the projection of the future onto the past is an N-dimensional subspace of the M-dimensional space to which the rows of $X_{L,2L-1}$ belong. Thus, the rank of matrix $X_{F \| P}$ is equal to the dimension of the subspace spanned by the projection of the future onto the past. Furthermore, the structure of this subspace (and of its matrix representation) determines the characteristics of the state space model (such as model order). Analogously, matrix $X_{F \perp P}$ determines the characteristics of the noise subspace.

31

The decomposition (3-4) can be carried out using a matrix operator referred to as a projector (Pease, 1965). Let S_1 and S_2 denote two orthogonal subspaces of S such that $S_1 \oplus S_2 = S$. A projector of S onto S_1 is a matrix $\mathbf{P}_1$ such that

$$(3\text{-}5a) \qquad \mathbf{P}_1 \underline{v}_1 = \underline{v}_1 \qquad\qquad \forall\, \underline{v}_1 \in S_1$$

$$(3\text{-}5b) \qquad \mathbf{P}_1 \underline{v}_2 = \underline{0} \qquad\qquad \forall\, \underline{v}_2 \in S_2$$

Projectors can be defined also as operating on row vectors, instead of on column vectors. The property which characterizes projectors is idempotency (that is, $\mathbf{P}$ is a projector if and only if $\mathbf{P}^2 = \mathbf{P}$).

Let $\mathcal{V}$ denote the M-dimensional subspace defined by the L rows of $X_{L,2L-1}$ (recall that $L \ll M$), and let $\mathbf{P}_-$ denote the MxM projector of $\mathcal{V}$ onto the subspace X^-. It follows that

$$(3\text{-}6) \qquad X_{L,2L-1}\mathbf{P}_- = X_{F_{\|}P}$$

Thus, availability of the projector $\mathbf{P}_-$ allows the decomposition of the future data matrix because $X_{F\perp P}$ can be determined from Equations (3-4) and (3-6). Projector $\mathbf{P}_-$ is determined from matrix $X_{0,L-1}$ as

$$(3\text{-}7) \qquad \mathbf{P}_- = X_{0,L-1}^{\dagger} X_{0,L-1} = X_{0,L-1}^{H}\left(X_{0,L-1}X_{0,L-1}^{H}\right)^{-1} X_{0,L-1}$$

This projector will decompose the future data matrix into the desired components. However, Equation (3-7) imposes a large computational burden, and furthermore, it effectively involves the calculation of the output correlation matrix sequence and of the inverse of a large matrix with correlation matrix sequence elements as its block elements. Fortunately, the QR decomposition

32

can be applied to determine the subspace decomposition (3-4) and also to determine the projector $P_\cdot$, if required. The QR decomposition is a computationally efficient and numerically robust approach to address this problem (Dongarra et al., 1979).

Consider now the block Hankel data matrix $X_{0,2L-1}$. This matrix is a 2JLxM block column matrix made up of a concatenation of the past and future Hankel matrices,

$$(3\text{-}8) \qquad X_{0,2L-1} = \begin{bmatrix} X_{0,L-1} \\ \text{------} \\ X_{L,2L-1} \end{bmatrix}$$

Now apply the Hermitian operator to a "normalized" matrix $X_{0,2L-1}$, and carry out a QR decomposition on this matrix (the normalization factor is required to avoid increase in dynamic range and to match the formulation which is based on the correlation matrix sequence). This results in

$$(3\text{-}9a) \qquad \frac{X_{0,2L-1}^H}{\sqrt{M}} = \frac{1}{\sqrt{M}} \begin{bmatrix} X_{0,L-1}^H & X_{L,2L-1}^H \end{bmatrix} = Q \begin{bmatrix} R^H \\ \text{---------} \\ O_{(M-2JL),2JL} \end{bmatrix}$$

$$(3\text{-}9b) \qquad \frac{X_{0,2L-1}^H}{\sqrt{M}} = \begin{bmatrix} Q_A & Q_B & Q_C \end{bmatrix} \begin{bmatrix} R_A^H & R_B^H \\ [0] & R_C^H \\ [0] & [0] \end{bmatrix}$$

Matrix Q is an MxM unitary matrix, submatrices Q_A and Q_B are dimensioned MxJL, and submatrix Q_C is dimensioned Mx(M-2JL). Matrix R^H is a 2JLx2JL upper-triangular matrix with rank equal to the rank of matrix $X_{0,2L-1}$. All the submatrices of R are

33

dimensioned $JL \times JL$, and submatrices R_A^H and R_C^H are also upper-triangular. Since matrix Q is unitary, the following relations are true:

$$(3-10) \qquad QQ^H = Q_AQ_A^H + Q_BQ_B^H + Q_CQ_C^H = I_M$$

$$(3-11) \qquad Q^HQ = \begin{bmatrix} Q_A^HQ_A & Q_A^HQ_B & Q_A^HQ_C \\ Q_B^HQ_A & Q_B^HQ_B & Q_B^HQ_C \\ Q_C^HQ_A & Q_C^HQ_B & Q_C^HQ_C \end{bmatrix} = \begin{bmatrix} I_{JL} & [0] & [0] \\ [0] & I_{JL} & [0] \\ [0] & [0] & I_{M-2JL} \end{bmatrix} = I_M$$

Consider now the conjugate transpose of Equation (3-9), after eliminating Q_C since it is multiplied by a zero-valued matrix; that is,

$$(3-12) \qquad \frac{X_{0,2L-1}}{\sqrt{M}} = \frac{1}{\sqrt{M}} \begin{bmatrix} X_{0,L-1} \\ \cdots \\ X_{L,2L-1} \end{bmatrix} = \begin{bmatrix} R_A & [0] \\ \cdots & \cdots \\ R_B & R_C \end{bmatrix} \begin{bmatrix} Q_A^H \\ \cdots \\ Q_B^H \end{bmatrix}$$

The following two equations are obtained immediately from the partitioning in Equation (3-12),

$$(3-13) \qquad \frac{X_{0,L-1}}{\sqrt{M}} = R_AQ_A^H$$

$$(3-14) \qquad \frac{X_{L,2L-1}}{\sqrt{M}} = R_BQ_A^H + R_CQ_B^H$$

Equation (3-13) is a QR decomposition of $X_{0,L-1}$ (recall that R_A is lower triangular), and Equation (3-14) is a subspace decomposition of $X_{L,2L-1}$. As shown next, (3-14) is the desired decomposition of

34

$X_{L,2L-1}$. The projector P_- is determined from Equations (3-7) and (3-13) as

$$(3-15) \qquad P_- = Q_A Q_A^H$$

It also follows that the projector of $\mathcal{V}$ onto $\mathcal{W}$ (the noise subspace) is

$$(3-16) \qquad P_W = Q_B Q_B^H$$

With Equations (3-11) and (3-15) it is easy to demonstrate that

$$(3-17) \qquad X_{L,2L-1} P_- = \sqrt{M} \left[R_B Q_A^H \right] = X_{F\parallel P}$$

Similarly, it follows from Equations (3-11) and (3-16) that

$$(3-18) \qquad X_{L,2L-1} P_W = \sqrt{M} \left[R_C Q_B^H \right] = X_{F\perp P}$$

This demonstrates that Equation (3-14) is the decomposition of the future onto the past and onto the noise orthogonal subspaces.

The information of the projection of the future onto the past is contained in matrix R_B. Specifically, the rank of R_B is equal to the order of the state space model representation for the future-to-past interface, and the column space of R_B is equal to the column space of the observability matrix for the state space model (Van Overschee and De Moor, 1993).

At this point in the development it is convenient to continue the decomposition of the Q and R matrices in Equation (3-12) in order to isolate as much as possible the structure of the orthogonal subspaces. To that end, consider Equation (3-12), and

35

carry out a further partitioning of the QR decomposition matrices
as follows (the dimensions of the matrices on the right-hand-side
of Equation (3-19a) are given in (3-19b) and (3-19c)):

$$
(3\text{-}19a) \quad \frac{X_{0,2L\text{-}1}}{\sqrt{M}} = \frac{1}{\sqrt{M}} \begin{bmatrix} X_{0,L\text{-}1} \\ \text{-----} \\ X_{L,2L\text{-}1} \end{bmatrix} = \begin{bmatrix} R_{11} & [0] & [0] & [0] \\ R_{21} & R_{22} & [0] & [0] \\ \text{-----} & \text{-----} & \text{-----} & \text{-----} \\ R_{31} & R_{32} & R_{33} & [0] \\ R_{41} & R_{42} & R_{43} & R_{44} \end{bmatrix} \begin{bmatrix} Q_1^H \\ Q_2^H \\ \text{----} \\ Q_3^H \\ Q_4^H \end{bmatrix}
$$

$$
(3\text{-}19b) \quad
\begin{array}{c}
 \\
J(L\text{-}1) \\
J \\
J \\
J(L\text{-}1)
\end{array}
\begin{array}{cccc}
J(L\text{-}1) & J & J & J(L\text{-}1) \\
\begin{bmatrix} R_{11} & [0] & [0] & [0] \\ R_{21} & R_{22} & [0] & [0] \\ \text{-----} & \text{-----} & \text{-----} & \text{-----} \\ R_{31} & R_{32} & R_{33} & [0] \\ R_{41} & R_{42} & R_{43} & R_{44} \end{bmatrix}
\end{array}
$$

$$
(3\text{-}19c) \quad
\begin{array}{c}
 \\
J(L\text{-}1) \\
J \\
 \\
J \\
J(L\text{-}1)
\end{array}
\begin{array}{c}
M \\
\begin{bmatrix} Q_1^H \\ Q_2^H \\ \text{----} \\ Q_3^H \\ Q_4^H \end{bmatrix}
\end{array}
$$

From Equations (3-12) and (3-19) it follows that the JLxJL matrices
R_B and R_C are defined with the following partitions:

$$
(3\text{-}20) \quad R_B = \begin{bmatrix} R_{31} & R_{32} \\ R_{41} & R_{42} \end{bmatrix}
$$

$$
(3\text{-}21) \quad R_C = \begin{bmatrix} R_{33} & [0] \\ R_{43} & R_{44} \end{bmatrix}
$$

36

Refer to the partitioning in Equation (3-19) and define two other partitioned matrices as

(3-22) $\quad R_D = [\ R_{41}\ \ R_{42}\ \ R_{43}\ ]$

(3-23) $\quad R_E = R_{44}$

Matrix R_D is $J(L-1) \times J(L+1)$, and matrix R_E is $J(L-1) \times J(L-1)$. Now carry out two QSVDs on these matrix pairs as detailed in Appendix B. One QSVD is applied to the matrix pair R_B and R_C to obtain

(3-24) $\quad R_B^H = U_L S_L Y_L^H$

(3-25) $\quad R_C^H = V_L T_L Y_L^H$

The second QSVD is carried out on the matrix pair R_D and R_E, and results in

(3-26) $\quad R_D^H = U_{L-1} S_{L-1} Y_{L-1}^H$

(3-27) $\quad R_E^H = V_{L-1} T_{L-1} Y_{L-1}^H$

In these two QSVDs, matrices U_{L-1}, U_L, V_{L-1}, and V_L are unitary, and matrices Y_{L-1} and Y_L are square and non-singular (Appendix B). Also, the subscripts (L or L-1) correspond to the term index of an associated observability matrix defined as in Equation (2-16). That is, the following two results are true:

- The column space of matrix R_D is the same as the column space of O_{L-1}. This result follows from two facts: first, for $L-1 \geq N$ the observability matrix maps the state space onto the output space; and second, the

37

decomposition on Equation (3-19) indicates a decomposition of the form (3-14) for the block Hankel matrix $X_{L+1,2L-1}$, which consists of the last L-1 block rows of $X_{0,2L-1}$ in Equation (3-19).

- The column space of matrix R_B is the same as the column space of O_L. As in the preceding argument, this follows from the mapping property of the observability matrix and from the decomposition of matrix $X_{L,2L-1}$ in Equation (3-14).

Consider the cases where the matrix pairs (R_B, R_C) and (R_D, R_E) form concatenated matrices of full rank (see Appendix B), which are the most likely cases in practical situations involving random data. In those cases matrix S_{L-1} is rectangular with 2J more rows than columns, and is zero except possibly along the main diagonal. The elements along the main diagonal of S_{L-1} are real-valued, with value bound between unity and zero, and arranged in order of decreasing magnitude. Matrices S_L, T_{L-1}, and T_L are square and diagonal. The diagonal elements of S_L are real-valued, with value bound between unity and zero, and arranged in order of decreasing magnitude also. The diagonal elements of both T_{L-1} and T_L are also real-valued and with value bound between unity and zero. However, the diagonal elements of these two matrices are arranged in order of decreasing magnitude. In pairs, the diagonal elements of S_{L-1} and T_{L-1} are referred to as singular value pairs of matrices R_D and R_E. Likewise, the diagonal elements of S_L and T_L are the singular value pairs of matrices R_B and R_C.

The value of the diagonal elements of matrices S_{L-1} and S_L are indicative of model order. In fact, when the data is the output of a system of order N, only the first N diagonal entries are non-zero in matrices S_{L-1} and S_L (model order determination is

38

discussed further in Section 3.2). As discussed in Appendix B, for every zero-valued diagonal entry in matrix S_L there is a corresponding unity-valued diagonal element in matrix T_L. The same relationship is true for matrices S_{L-1} and T_{L-1}. Thus, for an N-th order model the two pairs of $S_{(\cdot)}$ and $T_{(\cdot)}$ matrices have a natural partition along the main diagonal corresponding to the first N entries. Specifically,

$$(3\text{-}28) \qquad S_L = \begin{bmatrix} S_L^{(1)} & [0] \\ [0] & S_L^{(2)} \end{bmatrix} = \begin{bmatrix} S_L^{(1)} & [0] \\ [0] & O_{JL\text{-}N} \end{bmatrix}$$

$$(3\text{-}29) \qquad S_{L\text{-}1} = \begin{bmatrix} S_{L\text{-}1}^{(1)} & [0] \\ [0] & S_{L\text{-}1}^{(2)} \end{bmatrix} = \begin{bmatrix} S_{L\text{-}1}^{(1)} & [0] \\ [0] & O_{JL+J\text{-}N,JL\text{-}J\text{-}N} \end{bmatrix}$$

$$(3\text{-}30) \qquad T_L = \begin{bmatrix} T_L^{(1)} & [0] \\ [0] & T_L^{(2)} \end{bmatrix} = \begin{bmatrix} T_L^{(1)} & [0] \\ [0] & I_{JL\text{-}J\text{-}N} \end{bmatrix}$$

$$(3\text{-}31) \qquad T_{L\text{-}1} = \begin{bmatrix} T_{L\text{-}1}^{(1)} & [0] \\ [0] & T_{L\text{-}1}^{(2)} \end{bmatrix} = \begin{bmatrix} T_{L\text{-}1}^{(1)} & [0] \\ [0] & I_{JL\text{-}N} \end{bmatrix}$$

Now define block column partitions in matrices $U_{(\cdot)}$, $V_{(\cdot)}$, and $Y_{(\cdot)}$ to correspond with the partitions in Equations (3-28)-(3-31). This results in

$$(3\text{-}32) \qquad U_L = \begin{bmatrix} U_L^{(1)} & U_L^{(2)} \end{bmatrix} \qquad\qquad U_{L\text{-}1} = \begin{bmatrix} U_{L\text{-}1}^{(1)} & U_{L\text{-}1}^{(2)} \end{bmatrix}$$

$$(3\text{-}33) \qquad V_L = \begin{bmatrix} V_L^{(1)} & V_L^{(2)} \end{bmatrix} \qquad\qquad V_{L\text{-}1} = \begin{bmatrix} V_{L\text{-}1}^{(1)} & V_{L\text{-}1}^{(2)} \end{bmatrix}$$

$$(3\text{-}34) \qquad Y_L = \begin{bmatrix} Y_L^{(1)} & Y_L^{(2)} \end{bmatrix} \qquad\qquad Y_{L\text{-}1} = \begin{bmatrix} Y_{L\text{-}1}^{(1)} & Y_{L\text{-}1}^{(2)} \end{bmatrix}$$

All submatrices with superscript 1 have N columns. Partitioning of the $S_{(\cdot)}$, $T_{(\cdot)}$, $U_{(\cdot)}$, $V_{(\cdot)}$, and $Y_{(\cdot)}$ matrices does not provide a theoretical advantage or an enhanced insight; it does, however, reduce the computational burden.

The two QSVDs were introduced to extract the structure and subspace information available in the R matrix (and submatrices) of the QR decomposition. Substitution of the QSVD results into the corresponding partitions in Equations (3-12) and (3-19) allows appreciation of this structure. Carrying this out leads to the following expressions:

$$
(3\text{-}35) \qquad \frac{X_{L,2L-1}}{\sqrt{M}} = [\; R_B \;\vdots\; R_C \;] \begin{bmatrix} Q_A^H \\ \cdots\cdots \\ Q_B^H \end{bmatrix} = Y_L [\; S_L U_L^H \;\vdots\; T_L V_L^H \;] \begin{bmatrix} Q_A^H \\ \cdots\cdots \\ Q_B^H \end{bmatrix}
$$

$$
(3\text{-}36) \qquad \frac{X_{L+1,2L-1}}{\sqrt{M}} = [\; R_D \;\; R_E \;] \begin{bmatrix} Q_A^H \\ Q_B^H \end{bmatrix} = Y_{L-1} [\; S_{L-1} U_{L-1}^H \;\; T_{L-1} V_{L-1}^H \;] \begin{bmatrix} Q_A^H \\ Q_B^H \end{bmatrix}
$$

Both of these equations exhibit a subspace decomposition of the respective block Hankel data matrix, and in each equation the information of the structure of the two orthogonal subspaces is contained in the partitioned matrix involving the $S_{(\cdot)}$ and $T_{(\cdot)}$ matrices. In Equation (3-36) the partitions in the matrices are not emphasized because the dimensions of the individual partitions are not compatible (as they are in Equation (3-35)). Of course, overall matrix dimensions are compatible.

Given the decompositions in Equations (3-35) and (3-36), it remains to develop the procedure that relates these decompositions to the innovations model parameter matrices. This is done using orthogonal projections in random vector spaces.

40

Consider the orthogonal projection of the future block vector X_F onto the past block vector X_P (recall Equations (2-24) and (2-25)). This projection is the minimum variance estimate of X_F given X_P, which is also the conditional mean of X_F given X_P. That is,

$$(3-37) \qquad \hat{X}_F = E[X_F | X_P] = E[X_F X_P^H] (E[X_P X_P^H])^{-1} X_P$$

Using Equations (2-26) and (2-28) this can be expressed compactly in terms of the observability and reversed dual controllability matrices as

$$(3-38) \qquad \hat{X}_F = \overrightarrow{\mathcal{H}}_{L,L} \mathcal{R}_{P:L,L}^{-1} X_P = O_L \mathcal{B}_L \mathcal{R}_{P:L,L}^{-1} X_P$$

Suppose a minimum variance estimate is sought for each one of the columns of the data block Hankel matrix $X_{L,2L-1}$, which represents the future. Then, concatenating M such estimate equations into a single matrix estimate equation leads to

$$(3-39) \qquad \hat{X}_{L,2L-1} = O_L \mathcal{B}_L \mathcal{R}_{P:L,L}^{-1} X_{0,L-1} = O_L Z_L$$

The $N \times M$ matrix Z_L is very important, and deserves to be defined directly, as in Equation (3-40) next,

$$(3-40) \qquad Z_L = \mathcal{B}_L \mathcal{R}_{P:L,L}^{-1} X_{0,L-1}$$

Equation (3-39) states that the minimum variance estimate of the output (the columns of matrix $X_{L,2L-1}$) is a linear function of the columns of matrix Z_L. Recall that the observability matrix maps the state vector into the output, and that the states of a Kalman filter are minimum variance estimates of the states of the linear system for which the filter is designed. Thus, the columns of Z_L

41

are states of a Kalman filter for the system to be identified. This result is instrumental to the algorithm.

Equations (3-39) and (3-40) involve correlation matrices, which is undesired because of the computational burden associated with their calculation and also because of the increase in numerical precision (dynamic range) associated with computations involving correlation matrices. It is possible to convert these equations to "square root" form by substituting estimates of the correlation matrices calculated using channel output data. For sufficiently large values of M the reversed stochastic Hankel matrix and the past block correlation matrix are approximated effectively by the biased correlation matrix estimators using the channel output data, and it is simple to demonstrate that such estimates can be represented in terms of the output data block Hankel matrices. Specifically, for large M,

$$(3-41) \qquad \mathscr{R}_{P:L,L} = \frac{1}{M} X_{0,L-1} X_{0,L-1}^H = R_A R_A^H$$

$$(3-42) \qquad \overrightarrow{\mathscr{H}}_{L,L} = \frac{1}{M} X_{L,2L-1} X_{0,L-1}^H = Y_L S_L U_L^H R_A^H = \mathcal{Q}\mathscr{B}_L$$

where Equations (3-13) and (3-35) have been applied. Equation (3-42) suggests that the reversed stochastic Hankel matrix can be factorized, as in deterministic realization problems (Zeiger and McEwen, 1974), to obtain the observability and the reversed dual controllability matrices,

$$(3-43) \qquad \mathcal{Q}_L = Y_L S_L^{1/2} = Y_L^{(1)} \left(S_L^{(1)}\right)^{1/2}$$

$$(3-44) \qquad \mathscr{B}_L = S_L^{1/2} U_L^H R_A^H = \left(S_L^{(1)}\right)^{1/2} \left(U_L^{(1)}\right)^H R_A^H$$

42

Now the matrix of Kalman filter states, Z_L, can be determined as (using Equations (3-13), (3-40), (3-41), and (3-44)),

$$(3-45) \qquad Z_L = S_L^{1/2} U_L^H Q_A^H = \left(S_L^{(1)}\right)^{1/2} \left(U_L^{(1)}\right)^H Q_A^H = \left(S_L^{(1)}\right)^{1/2} \left(U_L^{(1)}\right)^H \begin{bmatrix} Q_1^H \\ Q_2^H \end{bmatrix}$$

This is a key result of the algorithm.

In order to calculate the system parameter matrices F, Γ, and H it is necessary to determine two additional filter state matrices: Z_{L+1} and W_L. Matrix Z_{L+1} can be thought of as a "shifted" version of matrix Z_L; in fact, the columns of matrix Z_{L+1} are obtained from the columns of matrix Z_L via a Kalman filter (or an innovations model, Equation (2-34)). Determination of matrix Z_{L+1} requires steps identical to those in the derivation of matrix Z_L. The main difference is that the matrices involved correspond to the (L+1)-term observability and reversed dual controllability matrices. And consequently, the results of the QSVD of matrices R_D and R_E are utilized. Repeating the steps in the derivation of matrix Z_L leads to the following result:

$$(3-46) \qquad Z_{L+1} = \left(S_L^{(1)}\right)^{-1/2} \left(\underline{Y}_L^{(1)}\right)^\dagger Y_{L-1}^{(1)} S_{L-1}^{(1)} \left(U_{L-1}^{(1)}\right)^H \begin{bmatrix} Q_1^H \\ Q_2^H \\ Q_3^H \end{bmatrix}$$

where the underbar denotes that matrix $\underline{Y}_L^{(1)}$ is obtained from matrix $Y_L^{(1)}$ by deleting the last block row.

If the procedure to determine matrix Z_L is followed departing from the orthogonal projection of the past block vector $\underline{X}_P$ onto the future block vector $\underline{X}_F$, then the resulting matrix, denoted as W_L, is a matrix of Kalman filter states for the backward Kalman filter

43

(the Kalman filter for the backward system). The form of this backward filter states matrix is

$$(3\text{-}47) \quad W_L = \left(S_L^{(1)}\right)^{1/2} \left[\; S_L^{(1)} \left(U_L^{(1)}\right)^H \quad T_L^{(1)} \left(V_L^{(1)}\right)^H \; \right] \begin{bmatrix} Q_1^H \\ Q_2^H \\ \text{----} \\ Q_3^H \\ Q_4^H \end{bmatrix}$$

The main difference between this expression and the expression for matrix Z_L, Equation (3-45), is that Equation (3-47) includes matrices T_L and V_L directly.

Given the Kalman states matrices Z_L, Z_{L+1}, and W_L, the system parameter matrices F, Γ, and H can be determined as least-squares solutions to linear systems of equations in noise. This is a result of the relationship with these matrices and the forward and backward Kalman filter for the channel output sequence $\{\underline{x}(n)\}$. The procedure is described next.

Since the columns of the matrices Z_L and Z_{L+1} are Kalman filter states, it follows that

$$(3\text{-}48) \quad Z_{L+1} = FZ_L + Z_u$$

where Z_u is a matrix of residuals orthogonal to Z_L (Kalman filter residuals are orthogonal to the state estimates). A least-squares estimate of F is obtained from Equations (3-45), (3-46), and (3-48) as

$$(3\text{-}49a) \quad F = Z_{L+1} Z_L^\dagger = Z_{L+1} Z_L^H \left(Z_L Z_L^H\right)^{-1}$$

44

$$(3\text{-}49b) \qquad F = \left(S_L^{(1)}\right)^{-1/2} \left(Y_L^{(1)}\right)^{\dagger} Y_{L\text{-}1}^{(1)} S_{L\text{-}1}^{(1)} \left(U_{L\text{-}1}^{(1)}\right)^{H} U_L^{(1)} \left(S_L^{(1)}\right)^{-1/2}$$

Alternative formulas for matrix F can be defined based on several relationships (observability; controllability; backward model; etc.) that are valid for the system matrix. Each distinct formula presents different numerical precision and computational requirements. This issue defines an important set of trade-offs for investigation in Phase II.

The output equation for the Kalman filter leads to the following matrix relation,

$$(3\text{-}50) \qquad X_L = H^H Z_L + Z_w$$

where matrix X_L is the first block row of matrix $X_{L,2L\text{-}1}$, and Z_w is a matrix of residuals orthogonal to Z_L. From the definition of X_L and Equations (3-12) and (3-19),

$$(3\text{-}51) \qquad X_L = \left[\, \underline{x}(L) \;\; \underline{x}(L+1) \;\; \cdots \;\; \underline{x}(L+M\text{-}1) \,\right] = \left[\, R_{31} \;\; R_{32} \;\; R_{33} \;\; [0] \,\right] \begin{bmatrix} Q_1^H \\ Q_2^H \\ \cdots \\ Q_3^H \\ Q_4^H \end{bmatrix}$$

From Equations (3-45), (3-50), and (3-51) the least-squares estimate of H^H is

$$(3\text{-}52a) \qquad H^H = X_L Z_L^{\dagger} = X_L Z_L^H \left(Z_L Z_L^H\right)^{-1}$$

$$(3\text{-}52b) \qquad H^H = \left[\, R_{31} \;\; R_{32} \,\right] U_L^{(1)} \left(S_L^{(1)}\right)^{-1/2}$$

45

This is a simple expression and it involves matrices of relatively low dimensionality. An alternative derivation for Equation (3-52b) departs from Equation (3-43) and capitalizes on the fact that matrix H^H occupies the first J rows of the observability matrix.

A least-squares estimate for Γ is obtained in a manner analogous to the solution for H^H obtained above. This is based on the fact that Γ^H is the output measurement matrix for the backward system. Thus, the output equation for the backward Kalman filter leads to the following matrix relation,

$$(3\text{-}53) \qquad X_{L\text{-}1} = \Gamma^H W_L + Z_v$$

where matrix $X_{L\text{-}1}$ is the last block row of matrix $X_{0,L\text{-}1}$, and Z_v is a matrix of residuals orthogonal to W_L. From the definition of $X_{L\text{-}1}$ and Equations (3-12) and (3-19),

$$(3\text{-}54) \qquad X_{L\text{-}1} = \begin{bmatrix} x(L\text{-}1) & x(L) & \cdots & x(L+M\text{-}2) \end{bmatrix} = \begin{bmatrix} R_{21} & R_{22} & [0] & [0] \end{bmatrix} \begin{bmatrix} Q_1^H \\ Q_2^H \\ \text{-}\text{-}\text{-}\text{-} \\ Q_3^H \\ Q_4^H \end{bmatrix}$$

Then, based on Equations (3-47), (3-53), and (3-54), the least-squares estimate of Γ^H is obtained as

$$(3\text{-}55a) \qquad \Gamma^H = X_{L\text{-}1} W_L^\dagger = X_{L\text{-}1} W_L^H \left(W_L W_L^H \right)^{\text{-}1}$$

$$(3\text{-}55b) \qquad \Gamma^H = \begin{bmatrix} R_{21} & R_{22} \end{bmatrix} U_L^{(1)} \left(S_L^{(1)} \right)^{1/2}$$

This expression is analogous to Equation (3-52). Just as in the case for H^H, an alternative derivation of Equation (3-55b) is

46

possible based on the fact that matrix Γ occupies the last J columns of the reversed dual controllability matrix, Equation (3-44).

Alternative formulas can be defined also for matrices H and Γ based on the various system relationships that involve these matrices. This also constitutes an important set of trade-offs for Phase II.

Notice that the $Q_{(\cdot)}$ matrices do not appear in the final expressions for the matrix parameters. The QR decomposition is fundamental to the algorithm, but only the $R_{(\cdot)}$ matrices have to be calculated and stored. This is a very important feature of the algorithm because one dimension of $Q_{(\cdot)}$ is very large (M), and the manipulation of these matrices would involve significant storage and computational requirements.

Determination of the remaining matrix parameters for the innovations model (2-29) is described next. Consider first the steady-state correlation matrix of the innovations model state, Π. This correlation matrix is equal to the correlation matrix of the Kalman filter state (Anderson and Moore, 1979). Therefore, a robust estimator for Π is based on the columns of matrix Z_L,

$$(3\text{-}56) \qquad \Pi = Z_L Z_L^H = S_L^{(1)}$$

It turns out that the backward filter states also lead to the same result,

$$(3\text{-}57) \qquad \Pi_b = W_L W_L^H = S_L^{(1)}$$

A system model such that the forward and backward correlation matrices are both diagonal and equal is said to be in <u>balanced</u>

47

coordinates (in the stochastic sense). Balanced coordinates allow effective model order selection and/or reduction.

The zero-lag output correlation matrix is obtained directly from the output sequence as

$$(3\text{-}58) \qquad \Lambda_0 = \frac{1}{N_T} \sum_{k=0}^{N_T-1} \underline{x}(k)\underline{x}^H(k)$$

$$(3\text{-}59) \qquad N_T = M + 2L - 1$$

here N_T is the total number of output data vectors (length of the output sequence) used in the algorithm. The innovations correlation matrix is obtained from Equation (2-30a),

$$(3\text{-}60) \qquad \Omega = \Lambda_0 - H^H \Pi H$$

Finally, the one-step prediction filter (Kalman) gain is obtained from Equation (2-32a) as

$$(3\text{-}61) \qquad K = [\Gamma - F\Pi H] \, \Omega^{-1} = [\Gamma - F\Pi H] \, [\Lambda_0 - H^H \Pi H]^{-1}$$

which completes the model parameter identification algorithm.

3.2 Model Order Determination

Model order determination is a necessary decision for any identification algorithm in applications where the true order of the system generating the channel output data is unknown, or where the true process generating the data may not belong to the model class adopted to represent the data. In the second case the model generated by the algorithm is a "representation model," as opposed to a "physical model" (a model based on accurate analyses of the

48

underlying physical processes). Determination of the model order is always a difficult problem, and the solution is rarely clear-cut. The Van Overschee-De Moor identification algorithm does have several strong features that lead to robust model order estimation. Principally, the algorithm identifies the model parameters of the innovations representation for the multichannel process in balanced coordinates. Model order determination is an important issue and deserves detailed analysis during Phase II.

The prime mechanism for model order selection in the algorithm is examination of the diagonal values of matrix S_L (recall that the diagonal elements of S_L are real-valued, non-negative, bounded by unity and zero, and are arranged in order of decreasing magnitude). As indicated earlier, the innovations model identified by the algorithm is in balanced coordinates (Moore, 1981), and the steady-state correlation matrices of the state of both the forward (Π) and backward (Π_b) innovations models are equal to matrix S_L. In a system representation in balanced coordinates the position of a state in the state vector is indicative of the importance of the contribution of that state to the output correlation sequence (the first state is equal in importance or more important than the second state; etc.), and the magnitude of the corresponding correlation matrix element is representative of the relative contribution of that state. Thus, an effective model order selection approach is to identify the negligible diagonal elements of matrix S_L, and select the model order to be equal to the number of non-negligible diagonal elements of S_L.

In most situations involving a finite amount of data, all the diagonal values in matrix S_L are different from zero. This is due to the fact that the subspace decomposition is imperfect with finite amounts of data because the measurement noise $\{\underline{w}(n)\}$ corrupts

49

the past output subspace, and vice versa. In such cases, model
order can be estimated by identifying jump discontinuities in the
magnitude of the diagonal values of S_L.

Other criteria can be examined to estimate model order.
Squaring the diagonal values of S_L emphasizes discontinuities, and
thus provides a good criterion also. The normalized running sum
of the diagonal values of S_L, and the normalized running sum of
the squared diagonal values of S_L, are two additional criteria for
model order selection.

In the absence of one or more jump discontinuities, external
information may be required, such as prior knowledge of the system
being modeled. Alternatively, a reasonable model order can be
selected, and various analyses can be carried out to reduce the
order of the model taking advantage of the features of a state
space realization in balanced coordinates.

Other considerations for model order determination arise in
the calculation of the QSVD for the matrix pair R_B and R_C. It
turns out that incorrect determination of the rank of a certain
matrix in the generation of the QSVD of the matrix pair R_B and R_C
can lead to major difficulties in the determination of model
order. As in Appendix B, let a $2JL \times JL$ matrix B denote the
concatenation of matrices R_B and R_C as (relevant equations from
Appendix B are repeated here for simplicity)

$$(3\text{-}62) \qquad B = \begin{bmatrix} R_B^H \\ \dots\dots \\ R_C^H \end{bmatrix}$$

Application of the QSVD to the matrix pair R_B and R_C leads to the
decomposition, expressed for the general case where $\text{rank}(B) < JL$,

50

$$\text{(3-63a)} \qquad B = \begin{bmatrix} U_L [\begin{array}{cc} S_L & O_{JL,JL-r(B)} \end{array}] \\ V_L [\begin{array}{cc} T_L & O_{JL,JL-r(B)} \end{array}] \end{bmatrix} Y_L^H$$

$$\text{(3-63b)} \qquad B^H = Y_L \begin{bmatrix} \begin{bmatrix} S_L^T \\ O_{JL-r(B),JL} \end{bmatrix} U_L^H & \begin{bmatrix} T_L^T \\ O_{JL-r(B),JL} \end{bmatrix} V_L^H \end{bmatrix}$$

where $r(B)$ denotes rank(B). As indicated in Section 3.1, the column space of matrix R_B is equal to the column space of O_L, and as is well known, the dimension of the column space of O_L is equal to the model order, N. Therefore, the decomposition of R_B in Equation (3-63) is indicative of model order. In fact, model order information is included in matrix S_L, since matrix Y_L is non-singular and matrix U_L is unitary. This is another interpretation of the order-determining properties of matrix S_L.

An important result associated with the decomposition (3-63) is Grassmann's dimension theorem. This theorem can be stated as

$$\text{(3-64)} \qquad \dim[\text{range}(R_B) \cap \text{range}(R_C)] = \text{rank}(R_B) + \text{rank}(R_C) - \text{rank}(B)$$

Another relevant result follows directly from first principles,

$$\text{(3-65)} \qquad \text{rank}(B) \geq \max[\text{rank}(R_B), \ \text{rank}(R_C)]$$

Joint consideration of Equations (3-64) and (3-65) indicates that the model order, N, satisfies the following constraint,

$$\text{(3-66)} \qquad N \leq \text{rank}(B) \leq JL$$

This equation implies that under-estimation of the rank of matrix B in the process of generating the QSVD for the matrix pair R_B and

51

R_C forces the dimension of the model to be a small value, whereas over-estimation of the rank of matrix B drives the dimension of the model towards large values. In the approach adopted herein the second condition is preferable because the innovations model parameters are estimated in balanced coordinates, and model order reduction is robust and effective in such cases, as discussed above. Therefore, in the generation of the QSVD for the matrix pair R_B and R_C it is preferred to select the rank of matrix B (or the respective matrices in the other QSVDs) to be as large as possible. In most practical cases where noise is present, the rank is likely to assume its maximum value, JL.

4.0 INNOVATIONS SEQUENCE GENERATION

In the approach pursued in this program, an unknown system of the form (2-2) is modeled as an innovations representation (2-29). Thus, once the innovations model parameters have been identified, an optimal Kalman filter can be configured to generate the innovations sequence, $\{\varepsilon(n)\}$, for the likelihood ratio calculations. The approach described in this section is applied to the observation data under each of the two hypotheses.

Any one of several equivalent Kalman filter formulations can be applied to generate the innovations sequence. However, the one-step predictor formulation offers significant advantages in the context of the intended application (Anderson and Moore, 1979). Specifically, the one-step predictor formulation generates the innovations sequence and the filter state update with a simple structure in the case where the input and output noises are correlated ($S \neq [0]$ in Equation (2-5a)), and thus imposes less real-time computational requirements than other formulations. Also, the model identification algorithm generates the parameters for the innovations model. Thus, the one-step predictor formulation is adopted in this work. Strictly speaking, the terminology "one-step predictor" should be used hereafter, but use of the term "Kalman filter" is accepted universally. Both terms are used herein.

The steady-state one-step predictor formulation for the innovations model (2-29) is a linear, time-invariant system described by the following equations:

(4-1a) $\hat{\underline{a}}(n+1|n) = F\hat{\underline{a}}(n|n-1) + K\varepsilon(n)$ $n \geq n_0$

(4-1b) $\varepsilon(n) = \underline{x}(n) - \hat{\underline{x}}(n|n-1) = \underline{x}(n) - H^H\hat{\underline{a}}(n|n-1)$ $n \geq n_0$

53

(4-1c) $\widehat{\underline{\alpha}}(n_o|n_o\text{-}1) = \underline{0}$

Here $\widehat{\underline{\alpha}}(n\text{+}1|n)$ is the estimate of the innovations model state vector at time $n\text{+}1$ based on observation data up to time n, $\widehat{\underline{x}}(n|n\text{-}1)$ is the estimate of the observation vector at time n based on observation data up to time $n\text{-}1$, and $\underline{\varepsilon}(n)$ is the innovations associated with the observation $\underline{x}(n)$. Matrix K is the steady-state filter gain matrix. The filter initial condition is set equal to zero because the innovations model initial condition is zero, Equation (2-29c). A block diagram of the Kalman filter is presented in Figure 4-1, displaying the channel output vector as input, and the innovations sequence vector as output.
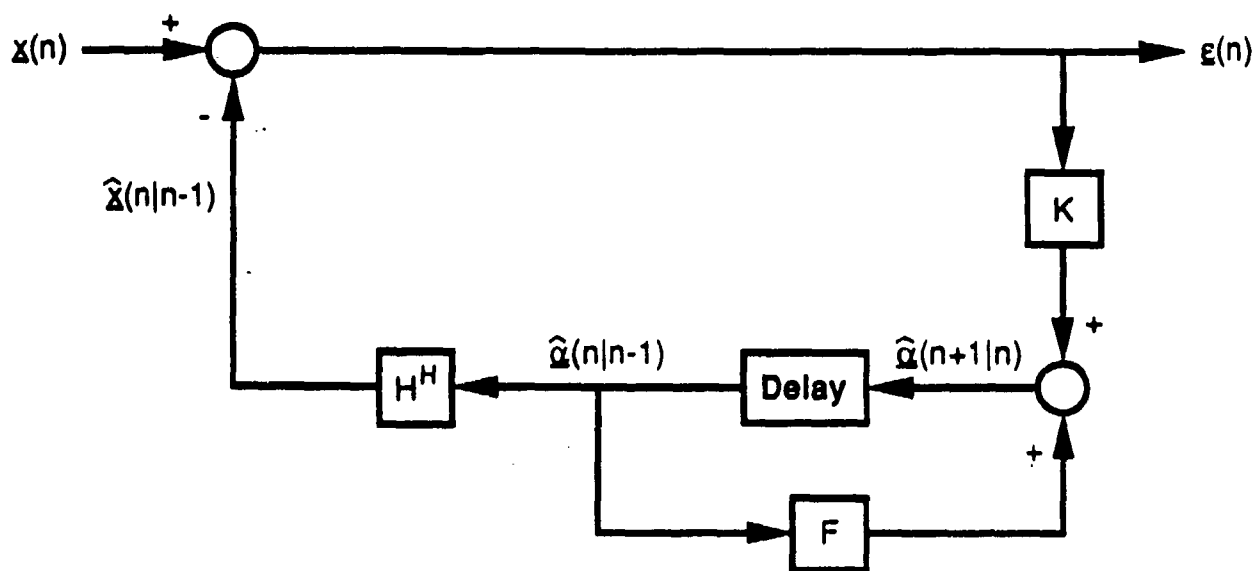
Figure 4-1. Kalman filter block diagram, emphasizing the
innovations sequence generation filter function.

The steady-state filter is an approximation to the optimal time-varying filter. If the channel output process is in steady-state, this approximation provides acceptable performance. Additionally, the steady-state filter provides a significant

54

reduction in the real-time computational requirements over the time-varying filter. In the cases where the channel output process is not in steady-state, filter performance is suboptimal, and the degree of loss of optimality needs to be ascertained. Such a determination is a subtask for Phase II. A related issue involves filter initialization transient effects. Since the steady-state filter gain is used, it may be necessary to neglect the first N_i filter outputs for each data batch. Determination of the value of N_i can be carried out via analysis and simulation, and is another subtask for Phase II.

Anderson and Moore (1979) show that the filter estimation error for an innovations model is zero at all times. That is,

$$(4-2) \qquad \hat{\alpha}(n+1|n) = \alpha(n+1)$$

Correspondingly, the filter estimation error correlation matrix is zero also. This can be inferred from the parallelism between the innovations model (2-29) and the filter representation (4-1). Thus, knowledge of the filter implies knowledge of the innovations model, and viceversa.

5.0 LIKELIHOOD RATIO DETECTION

A detection methodology for complex-valued multichannel Gaussian processes has been developed by Michels (1991) in the context of innovations-based detection. This approach has been generalized recently to include a class of non-Gaussian processes known as spherically-invariant random processes (SIRPs) and using linear estimators (Rangaswamy, Weiner, and Michels, 1993). Michels' methodology can be applied directly to the innovations sequence generated by the approach formulated herein. For brevity, only the likelihood ratio equation is presented here.

As discussed in Section 4.0, a Kalman filter (one-step predictor) is determined for each of the two hypotheses based on processing the multichannel data. The model order for the alternative hypothesis (H_1) filter is chosen to be larger than the model order for the null hypothesis (H_0) filter. For each hypothesis filter, denote the innovations sequence, Equation (4-1b), as

$$(5\text{-}1) \qquad \underline{\varepsilon}(n|H_i) = \underline{x}(n) - \underline{\hat{x}}(n|n\text{-}1;H_i) = \underline{x}(n) - H^H \underline{\hat{\underline{x}}}(n|n\text{-}1;H_i) \qquad i = 0, 1$$

The steady-state correlation matrix of the innovations is denoted as $\Omega(H_i)$, and is defined in Equation (3-60).

Let $\Theta(H_0, H_1)$ denote the multichannel likelihood ratio as defined by Michels (1991) for the Gaussian signal case. Then, the log-likelihood ratio (LLR) can be expressed as follows,

$$(5\text{-}2) \qquad \ln[\Theta(H_0, H_1)] = \sum_{n=n_0}^{N_T} \left[\ln\left[\frac{|\Omega(H_0)|}{|\Omega(H_1)|} \right] + \underline{\varepsilon}^H(n|H_0)\, \Omega^{-1}(H_0)\, \underline{\varepsilon}(n|H_0) \right.$$

$$\left. - \underline{\varepsilon}^H(n|H_1)\, \Omega^{-1}(H_1)\, \underline{\varepsilon}(n|H_1) \right]$$

56

The LLR is compared to a threshold, $\mathcal{T}$, which is calculated adaptively to maintain a constant false alarm rate (CFAR),

$$(5\text{-}3) \qquad \ln\left[\Theta(H_0, H_1)\right] = \begin{cases} \geq \mathcal{T} & \text{select } H_1 \\ < \mathcal{T} & \text{select } H_0 \end{cases}$$

A candidate CFAR approach with demonstrated good performance calculates the median of a set of the LLR values from a number of adjacent range cells (at the same azimuth) on both sides of the cell in question, and scales the calculated median value by a pre-determined constant to provide the desired false alarm rate (Metford and Haykin, 1985).

The LLR expression has to be modified if optimal time-varying filters are used instead of the steady-state filters. In such cases the modification is straightforward, and involves replacing the steady-state correlation matrices of the two innovations by their time-varying values.

Alternative expressions for the log-likelihood ratio can be generated based on factorization of the innovations correlation matrix and spatial whitening of the innovations process. This includes Cholesky factorization, LDU decomposition, and SVD. The first two techniques have been described by Michels (1991), and lead to simplified LLR expressions. The SVD technique is derived here.

Consider the steady-state innovations correlation matrices for each of the two hypotheses and carry out an SVD on each correlation matrix. This results in the following decompositions:

$$(5\text{-}4) \qquad \Omega(H_i) = V_i \Sigma_i V_i^H \qquad\qquad\qquad i = 0, 1$$

where matrix V_i is a JxJ unitary matrix, and Σ_i is a diagonal matrix with real-valued, positive elements arranged along the diagonal in decreasing order of magnitude (it is assumed herein that the correlation matrix of the innovations sequence has full rank). That is,

(5-5a)
$$\Sigma_i = \begin{bmatrix} \sigma_{i1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{i2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{iJ}^2 \end{bmatrix} \qquad i = 0, 1$$

(5-5b)
$$\sigma_{i1}^2 \geq \sigma_{i2}^2 \geq \ldots \geq \sigma_{iJ}^2 > 0 \qquad i = 0, 1$$

Since matrix V_i is unitary, the determinant and inverse functions of $\Omega(H_i)$ are obtained easily as

(5-6)
$$\Omega^{-1}(H_i) = V_i \Sigma_i^{-1} V_i^H \qquad i = 0, 1$$

(5-7)
$$|\Omega(H_i)| = \prod_{k=1}^{J} \sigma_{ik}^2 \qquad i = 0, 1$$

Now make a linear transformation on the innovations sequence using the unitary matrix V_i, to obtain

(5-8)
$$\underline{y}(n|H_i) = V_i^H \underline{g}(n|H_i) \qquad i = 0, 1$$

The transformed innovations sequence, $\underline{y}(n|H_i)$, is uncorrelated spatially as well as temporally (recall that $\underline{g}(n|H_i)$ is uncorrelated temporally), with correlation matrix Σ_i. Transformation of a J-dimensional vector by a unitary matrix rotates the vector in the J-dimensional space, but does not alter its magnitude. Thus, the

58

spatial whitening transformation does not alter the variance of the elements of the innovations vector.

Substituting Equations (5-4) through (5-8) into Equation (5-2) results in the following LLR expression

$$(5-9) \qquad \ln\!\left[\Theta(H_0, H_1)\right] = \sum_{n=n_o}^{N_T} \sum_{k=1}^{J} \left[\ln\!\left[\frac{\sigma_{0k}^2}{\sigma_{1k}^2}\right] + \frac{|v_k(n|H_0)|^2}{\sigma_{0k}^2} - \frac{|v_k(n|H_1)|^2}{\sigma_{1k}^2} \right]$$

where $v_k(n|H_i)$ denotes the kth element of $\underline{v}(n|H_i)$. This LLR is of the same form as the LLR derived by Michels (1991) for spatial whitening of the innovations using an LDU decomposition.

59

6.0 SOFTWARE SIMULATION

The identification and filtering algorithms described in the preceding sections have been programmed in FORTRAN 77 for Apple Macintosh processors. Support software for the validation and execution of the routines has been generated also. The support software includes signal generation routines, auxiliary routines for validation, and code for miscellaneous calculations. The identification algorithm makes use of the singular value and QR decompositions. Subroutines that implement these matrix operations for complex-valued matrices were obtained from versions of the LINPACK software package (Dongarra et al., 1979). Separate code was written and exercised to validate the LINPACK routines before incorporation into the main algorithm code. The signal generation code uses a Gaussian random number generator obtained from the text by Press et al. (1989). Sample realizations generated by this code were tested for whiteness and gaussianity.

6.1 Software Validation

Code validation was carried out in two steps. First, all subroutines and select segments of code were validated individually. Second, the complete package was validated using examples generated for that purpose. The examples consisted of system models with a simple structure so that the computer output could be predicted. Both real-valued and complex-valued examples were generated. One particular example used is the second-order system defined by the following matrix parameters (for a system model of the form (2-2)):

$$ F = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} $$

60

$$H^H = G = D^H = Q = C = I_2$$

This model was used to generate a random vector sequence to validate various aspects of the software. For example, defining matrix F with $f_{11} = f_{12} = f_{22} = 0$ and $f_{21} = 1$ generates an output vector sequence that consists of white noise in each output channel, but the two channels are correlated from one instant to the next (the correlation is due to the coupling induced by the non-zero (2,1) element of F). The output of the identification program should indicate a first-order model, with the first diagonal element of matrix S_L approximately equal to 0.7071, and low values for the remaining diagonal elements. This was the result obtained. Complex-valued test cases using this sample model were generated by letting F be a diagonal matrix with the desired complex-valued poles along the diagonal.

During validation and testing it was discovered that system poles along the real axis are more difficult to estimate, and that Equation (3-49) can produce biased results (over-estimation of the system poles) in some cases. This has been observed also by the research team at the Catholic University of Leuven, and one approach to mitigate this consists of utilizing alternative, more complex algorithms for the estimation of matrix F. Identification and detailed evaluation of these alternatives is an activity for Phase II of this program.

6.2 Analyses and Simulation Results

The software has been exercised with cases generated using multichannel AR models provided by the program monitor at RL, Dr. James H. Michels. These cases consist of signal only, clutter only, signal + noise, clutter + noise, and signal + clutter +

noise. In all cases the signal, clutter, and noise processes are statistically independent of each other.

Signal AR Model

The signal model is a complex-valued, two-input, two-output AR model of order 2 with the following matrix parameters,

$$\underline{y}_s(n) = -A_s^H(1)\underline{y}_s(n\text{-}1) - A_s^H(2)\underline{y}_s(n\text{-}2) + \underline{u}_s(n)$$

$$A_s^H(1) = \begin{bmatrix} 1.6290 - j\,1.4241 \times 10^{-7} & 1.3733 \times 10^{-5} + j\,3.8202 \times 10^{-13} \\ 1.3733 \times 10^{-5} + j\,3.8202 \times 10^{-13} & 1.6290 - j\,1.4241 \times 10^{-7} \end{bmatrix}$$

$$A_s^H(2) = \begin{bmatrix} 0.80996 - j\,1.4162 \times 10^{-7} & 1.5259 \times 10^{-5} - j\,9.0949 \times 10^{-13} \\ 1.5259 \times 10^{-5} - j\,9.0949 \times 10^{-13} & 0.80996 - j\,1.4162 \times 10^{-7} \end{bmatrix}$$

The input to the signal AR recursion, $\{\underline{u}_s(n)\}$, is a zero-mean, unit variance white noise sequence with a spatial correlation structure defined as

$$Q_s = \begin{bmatrix} 0.13038 & 0.12907 \\ 0.12907 & 0.13038 \end{bmatrix}$$

This two-input, two-output AR model corresponds to a fourth-order state space model in an innovations representation (as described in Appendix A), with poles at the following locations in the complex z-plane:

True Signal Model Poles: $-0.81451 \pm j\,0.38282$
$-0.81449 \pm j\,0.38281$

This AR system was defined by Michels to have a very high channel-to-channel correlation (~0.99), which indicates that a lower-order

62

model could represent the signal information. Specifically, a second-order state space model can represent the signal information well. Notice that the pole locations are almost repeated roots, which indicates that the two channels are almost identical. Therefore, given a high-level of channel-to-channel correlation, a reduced-order model should perform adequately.

The AR process $\{\underline{y}_s(n)\}$ is corrupted by a zero-mean, unit-variance white noise sequence $\{\underline{w}(n)\}$ to give the noise-corrupted channel output sequence as

$$\underline{x}_s(n) = \underline{y}_s(n) + \underline{w}(n)$$

For this noise model and the signal model given above, the signal-to-noise ratio (SNR) is approximately 3 dB.

Consider the problem of representing the AR signal in additive white noise with a state space model. The channel output noise, $\{\underline{w}(n)\}$, alters the parameters of the state space model designed for the AR signal $\{\underline{y}_s(n)\}$ only, but $\{\underline{x}_s(n)\}$ can be represented as the output of a state space model. That is, $\{\underline{y}_s(n)\}$ is represented as the output of an innovations model, but the model for $\{\underline{x}_s(n)\}$, which includes the additive noise $\{\underline{w}(n)\}$, is not an innovations model (there is an innovations model for $\{\underline{x}_s(n)\}$, but it is different from the innovations model for $\{\underline{y}_s(n)\}$). This is a manifestation of the well-known fact that an AR process corrupted by additive output white noise is no longer an AR process. In contrast, the state space model remains a valid representation of the signal even after the addition of a new noise source. This is one of the reasons why algorithms developed based on state space models are more robust than algorithms based on time series models.

Clutter AR Model

The clutter model is a complex-valued, two-input, two-output AR model of order 2 with the following matrix parameters,

$$\underline{y}_c(n) = - A_c^H(1)\underline{y}_c(n\text{-}1) - A_c^H(2)\underline{y}_c(n\text{-}2) + \underline{u}_c(n)$$

$$A_c^H(1) = \begin{bmatrix} -1.0430 & 0.0 \\ 0.0 & -1.0430 \end{bmatrix}$$

$$A_c^H(2) = \begin{bmatrix} 0.4900 & 0.0 \\ 0.0 & 0.4900 \end{bmatrix}$$

The input to the clutter AR recursion, $\{\underline{u}_c(n)\}$, is a zero-mean, unit variance white noise sequence with a spatial correlation structure defined as

$$Q_c = \begin{bmatrix} 1.5502 & 0.0 \\ 0.0 & 1.5502 \end{bmatrix}$$

This two-input, two-output AR model corresponds to a fourth-order state space model in an innovations representation (see Appendix A), with poles at the following locations in the complex Z-plane:

<u>True Clutter Model Poles</u>: $0.5215 \pm j\ 0.4669$

$$0.5215 \pm j\ 0.4669$$

The clutter AR coefficient values, the noise covariance values, and the diagonal structure of this AR system indicate that the two channels are uncorrelated. Thus, a fourth-order state space model can represent the clutter information well. Notice that the pole locations are repeated roots, which indicates that the two channels are identical.

The clutter AR process $\{\underline{y}_c(n)\}$ is corrupted by the zero-mean, unit-variance white noise sequence $\{\underline{w}(n)\}$ to give the noise-corrupted channel output sequence as

$$\underline{x}_c(n) = \underline{y}_c(n) + \underline{w}(n)$$

For this noise model and clutter model the clutter-to-noise ratio (CNR) is approximately 6 dB.

Selected Simulation Results

The identification and filtering software was exercised with the sequence $\{\underline{x}_s(n)\}$. Calculated values of the model order criterion parameters, the diagonal elements of matrix S_L, indicate that a second-order state space model is a good approximation to this system, as expected. Specifically, the values of the diagonal elements of matrix S_L are: {0.9804, 0.9776, 0.6712, 0.6360, 0.1215, 0.0912}, and the set of the square of these values is: {0.9612, 0.9557, 0.4505, 0.4045, 0.0148, 0.0083}. Examination of these two sets indicates that a second-order model is a good fit to the data. The next reasonable model order selection is four.

Based on the above discussions, model order 2 was selected for the AR signal in white noise. Figure 6-1 is a plot of the real part of the first element of a single realization of the innovations vector process, $\{\varepsilon_1(n)\}$, generated using a filter of order 2 (all plots herein are for single-realization cases). The filter parameters were identified using 2,214 output sequence vectors (corresponding to $L = 8$ and $M = 2,200$ in Equations (3-21) and (3-22)). Only 500 points are shown in the figure, but these points are representative of the sample process. The innovations sequence appears to be unbiased, with a calculated sample mean of

65

$$\bar{\xi}(n) = \begin{bmatrix} 0.0103 + j\,0.0074 \\ 0.0308 - j\,0.0050 \end{bmatrix}$$

Notice also the high degree of "whiteness" exhibited by the innovations. The imaginary part of the innovations behaves similarly. The autocorrelation function of the sequence in Figure 6-1 was estimated, and is shown in Figure 6-2. This figure clearly indicates the random (white) nature of the innovations, as expected. The zero-lag innovations correlation matrix identified by the software using Equation (3-60) is

$$\Omega = \begin{bmatrix} 1.5337 & 0.5908 + j\,0.0453 \\ 0.5908 - j\,0.0453 & 1.5832 \end{bmatrix}$$

This agrees very well with the sample correlation value of 1.536 indicated in Figure 6.2. Several simulation runs were made using multiple sample realizations of the same length and filter order two. A plot of the innovations correlation averaged over ten realizations looks very similar to Figure 6-2.
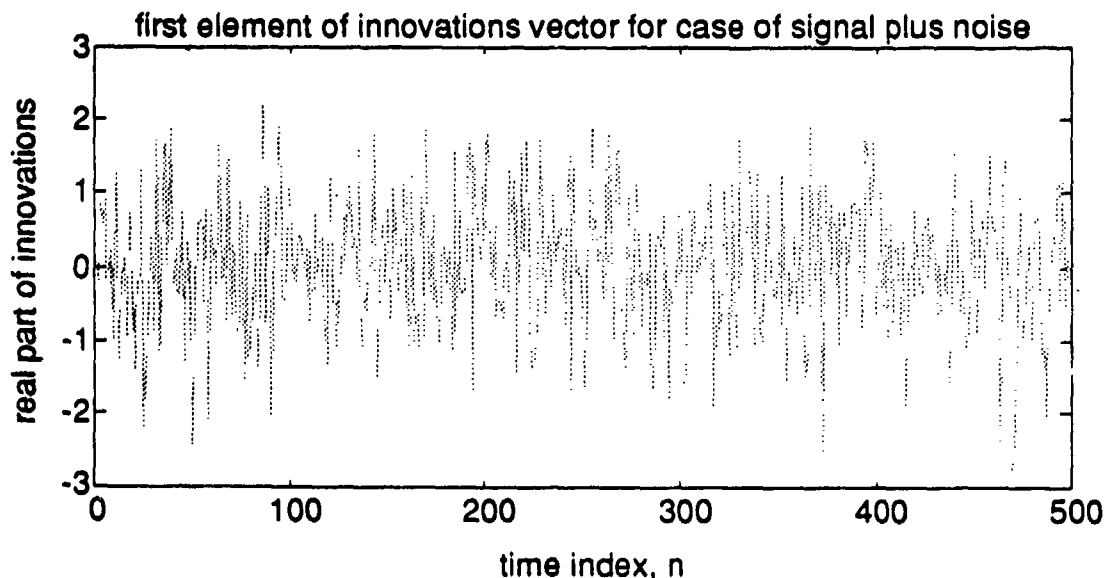

Figure 6-1.   Real part of the first element of innovations vector for the case of signal plus noise.

66
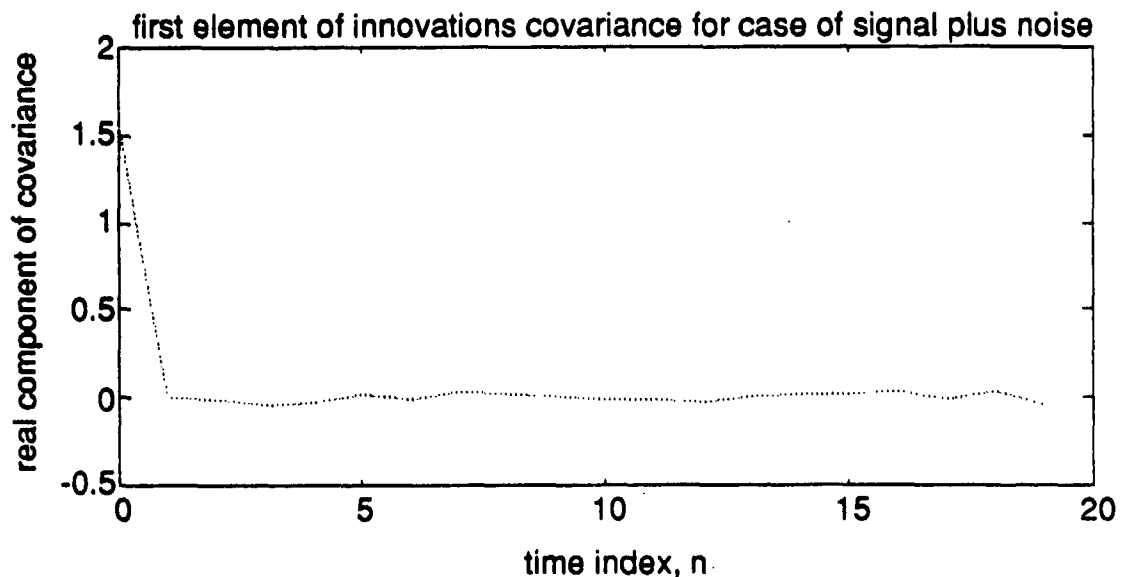
Figure 6-2. Real part of the auto-correlation function of the first element of the innovations sequence vector for the case of signal plus noise.

Identification algorithm performance can be assessed by examining the roots of the identified innovations model system matrix, F. The scatter plots in Figure 6-3, which correspond to results obtained for ten distinct realizations, illustrate the parameter identification capability of the algorithm. These scatter plots show the ten identified root pairs, all in close proximity to the true roots given above. The dashed lines in each plot intersect at the center of each plot ($-0.81 - j\ 0.38$ and $-0.81 + j\ 0.38$, respectively), and the centers are very close to the true root values ($-0.8145 \pm j\ 0.3828$). All the identified roots are at a distance less than 3% of the true values, and most are much closer than that.

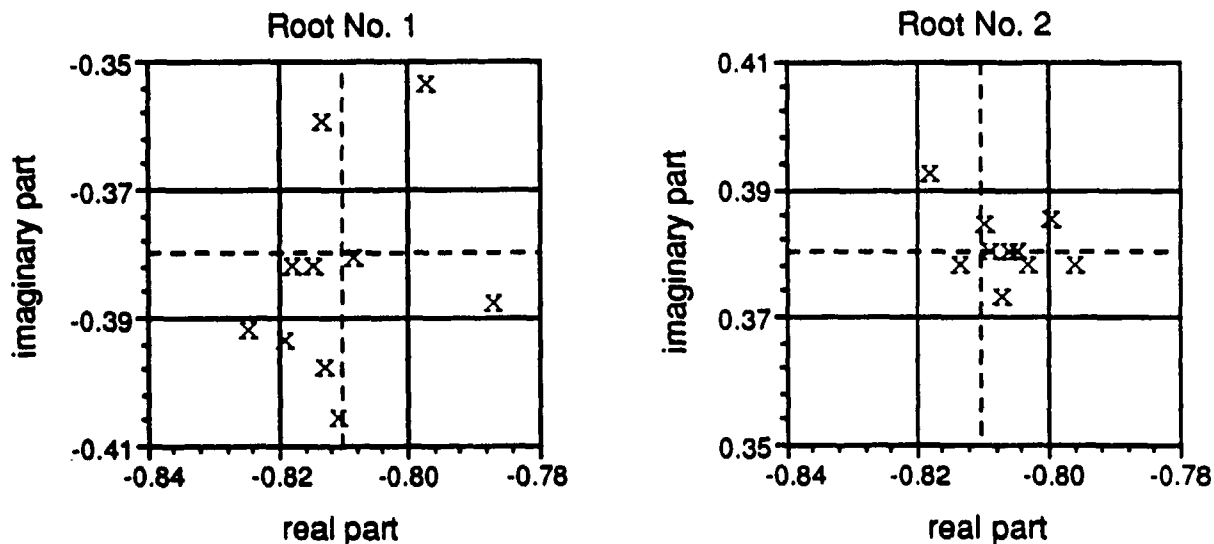Figure 6-3. Scatter plot of real and imaginary parts of identified model poles for ten distinct realizations of signal plus noise.

The software was used also to model and analyze the clutter plus noise sequence, $\{x_c(n)\}$. For this case at a CNR of 20 dB, the model order criterion parameters, the diagonal elements of the matrix S_L are: {0.8187, 0.7919, 0.3027, 0.2748, 0.1340, 0.1112}; and the set of the square of these values is: {0.6703, 0.6271, 0.0916, 0.0755, 0.0180, 0.0124}. This information, together with knowledge of the lack of channel correlation, indicates that a fourth-order state-space model is a good approximation to this system. Without prior knowledge regarding channel correlation, measures such as the percentage incremental power attributable to each additional singular value may result in improved model order estimates.

Model order four was selected for state space representation of the clutter AR process in additive white noise. Plots of the real and imaginary parts of the first element of a single realization of the innovations vector process, $\{\varepsilon_1(n)\}$, are presented

68

in Figure 6.4. These results were generated using a fourth-order filter for a case with 6 dB CNR. The filter parameters were identified using 2,214 output sequence vectors (corresponding to $L = 8$ and $M = 2,200$ in Equations (3-21) and (3-22)), as in the signal + noise case. Only 500 points are shown Figure 6.4, but these points are representative of the sample process. Both components (real and imaginary) of the sequence $\{\varepsilon_1(n)\}$ are unbiased, as indicated by the sample mean,

$$\bar{\varepsilon}(n) = \begin{bmatrix} -0.0101 - j\,0.0429 \\ 0.0538 + j\,0.0076 \end{bmatrix}$$

An estimate of the real and imaginary parts of the sample autocorrelation function of $\{\varepsilon_1(n)\}$ of Figure 6-4 is given in Figure 6-5. The real part has an impulse at lag $n = 0$ and is close to zero everywhere else, which is representative of a white innovations. The imaginary part exhibits low-amplitude oscillations about zero, as expected of a white innovations. The zero-lag innovations correlation matrix estimated using Equation (3-60) is

$$\Omega = \begin{bmatrix} 3.1114 & 0.0594 + j\,0.0248 \\ 0.0594 - j\,0.0248 & 3.3048 \end{bmatrix}$$

Element $(1,1)$ agrees with the sample correlation value of $3.106 + j\,0.0$ indicated in Figure 6-5. The behaviour of $\{\varepsilon_2(n)\}$ is similar.

Figure 6-6 presents scatter plots of the poles of the fourth-order system for ten realizations. The roots are clustered about the values of the true repeated roots $(0.5215 \pm j\,0.4669)$, which are indicated by the intersections of the dashed lines. The largest root estimation error is less than 8.2%. This error is larger than the worst error in the signal plus noise case, and is due to the greater difficulty in estimating faster modes.

69
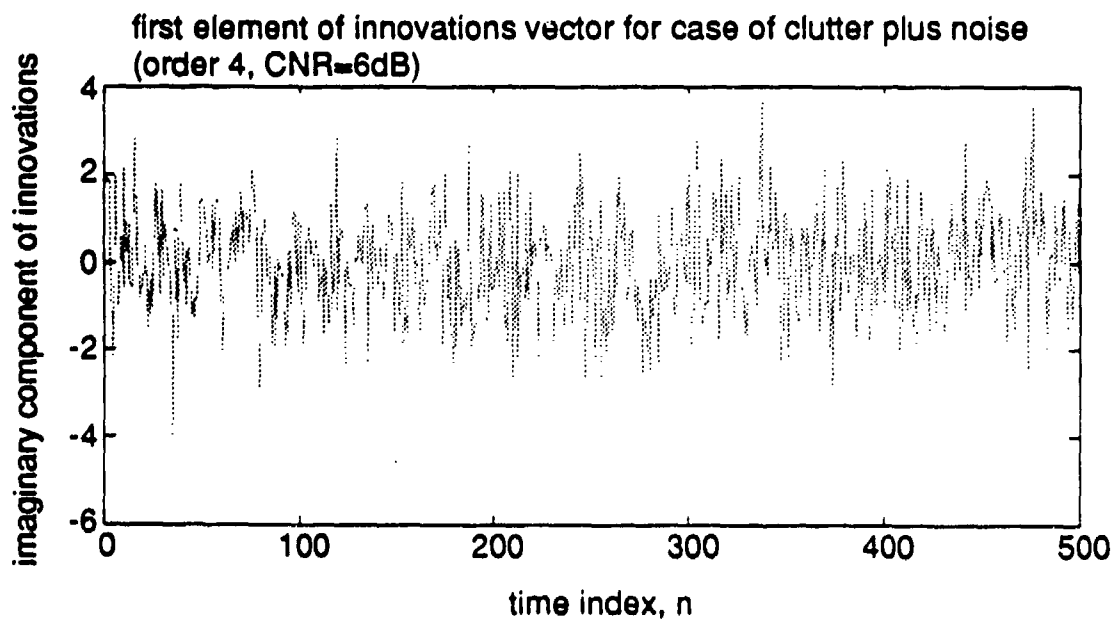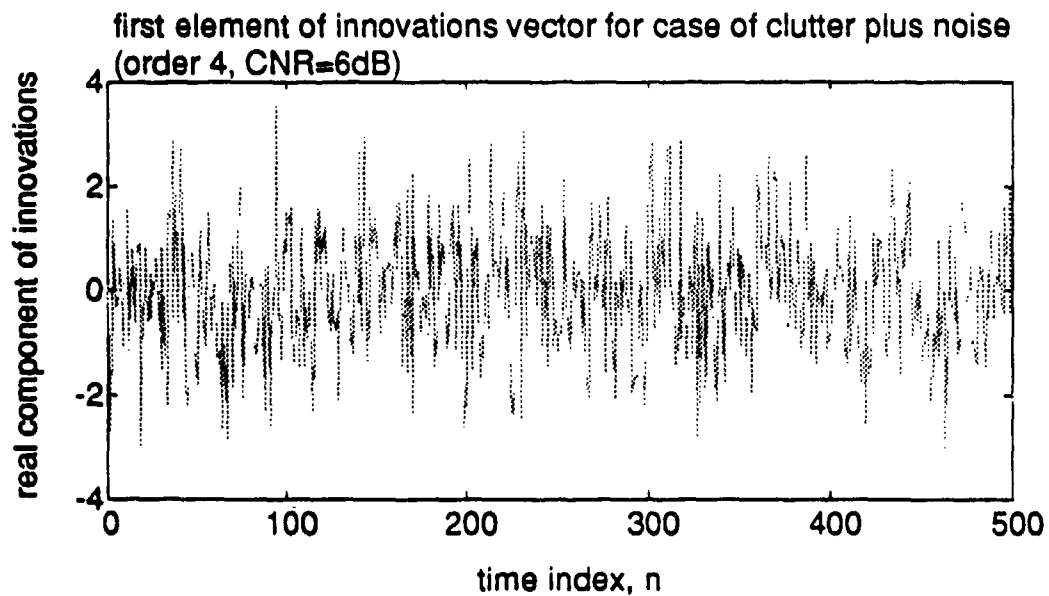
Figure 6-4. Real and imaginary parts of the first element of the innovations sequence vector for the case of clutter plus noise (CNR = 6 dB conditions).

70

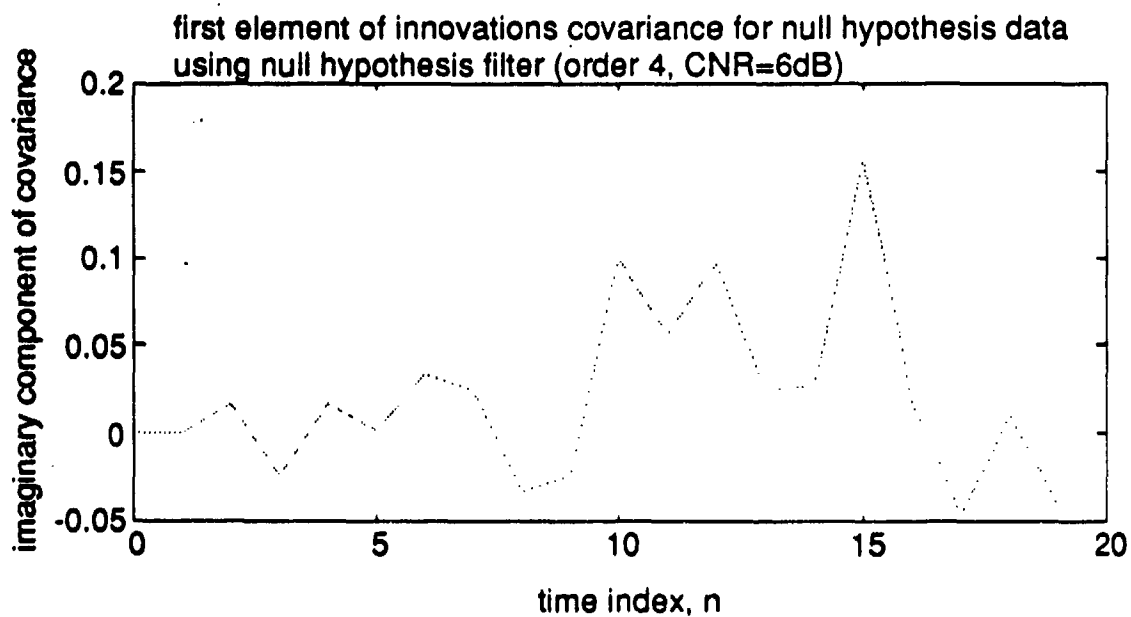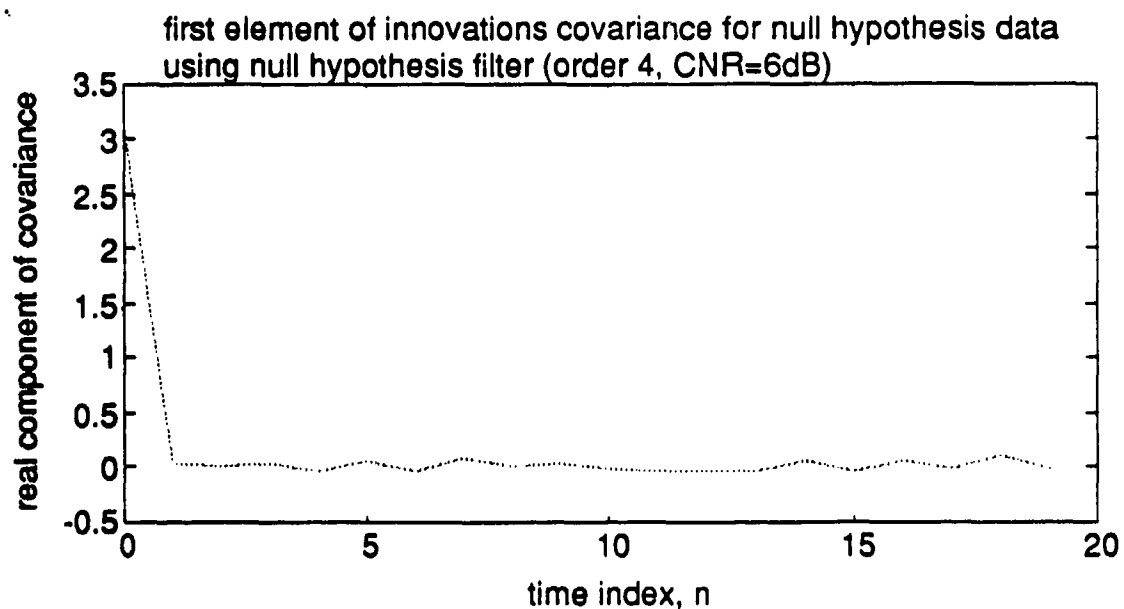Figure 6-5. Real and imaginary parts of the auto-correlation function of the first element of the innovations sequence vector for the case of clutter plus noise (CNR = 6 dB).
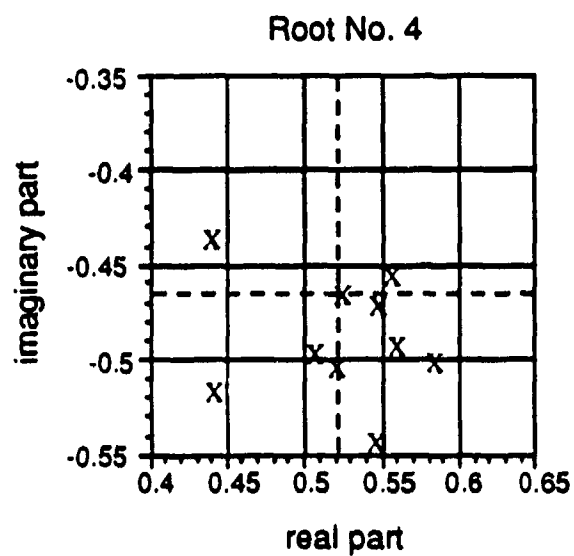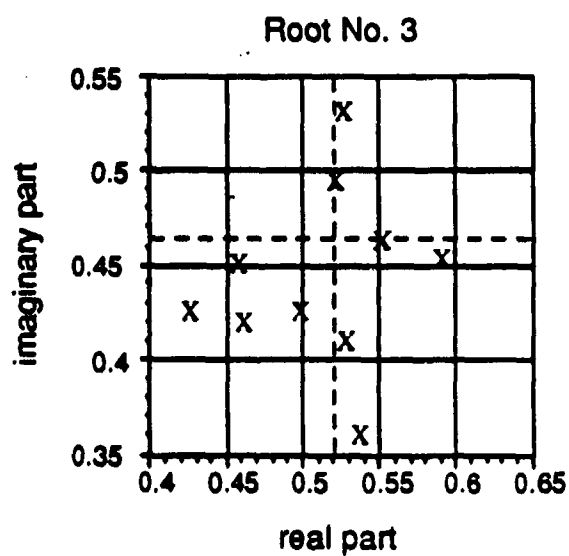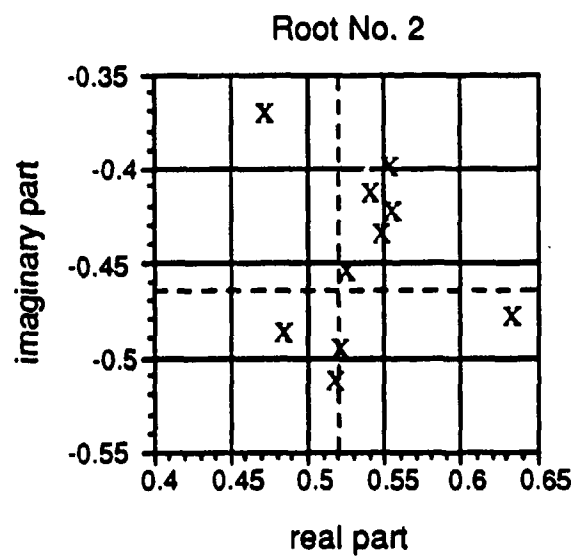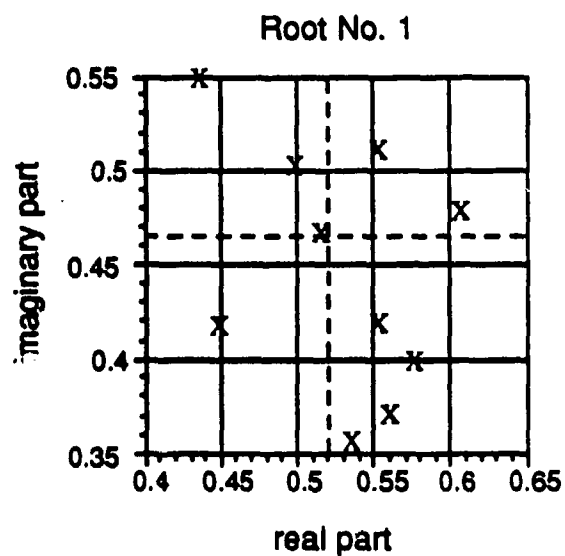
Figure 6-6. Scatter plot of real and imaginary parts of identified model poles for ten distinct realizations of clutter plus noise.

72

Similar biased behaviour has been observed in other estimation results (Michels, 1992b), as well as in detection performance results (Michels, 1992a) obtained using time series (AR) models. For the state-space approach pursued herein, unbiased estimates with reduced variance can be obtained by averaging several individual estimates and/or by increasing the complexity of the estimation algorithm.

Various simulations were carried out to obtain a first-order assessment of the discrimination capability of the innovations-based methodology using the SSC algorithm. One set of simulations involved designing a Kalman filter for each hypothesis, processing data corresponding to each of the two hypotheses using both filters, and analyzing the resulting four filter output sequences (two filters, and each filter processes data sets corresponding to each of the two hypotheses). These results are presented next. As before, all plots correspond to single-realization cases.

Consider first the case of processing data from each of the two hypotheses using a null hypothesis filter, corresponding to clutter + noise only. For this case the filter order is four, as mentioned earlier in the clutter plus noise model discussion. Results are presented herein for two sets of conditions: (a) SNR = 3 dB and CNR = 6 dB; and (b) SNR = 3 dB and CNR = 20 dB. For each set of conditions the procedure described next was followed.

- A realization of the clutter + noise process of duration $M = 2,200$ was generated and processed to design a fourth-order Kalman filter. The resulting filter is the filter for the null hypothesis (signal not present).

- The null hypothesis filter was applied to a clutter + noise process sequence of duration $M = 2,200$, and the

73

sample correlation matrix sequence of the filter output sequence was calculated. The real and imaginary parts of the (1,1) element of the resulting sample correlation matrix sequence are plotted in Figure 6-5 for CNR = 6 dB conditions, and Figure 6-7 for CNR = 20 dB conditions. Both sets of figures are representative of the auto-correlation of a white innovations sequence, as expected (both sets of figures show low-level energy content at the higher lags).

- The null hypothesis filter was applied to a combined signal + clutter + noise process sequence (alternative hypothesis case) of duration $M = 2,200$, and the sample correlation matrix sequence of the filter output sequence was calculated. In this case, however, the sequence is not a true innovations sequence because the filter is not optimal for this process. The real and imaginary parts of the (1,1) element of the resulting sample correlation matrix sequence are plotted in Figure 6-8 for CNR = 6 dB conditions, and Figure 6-9 for CNR = 20 dB conditions. Both of these figures show a marked deviation from the expected auto-correlation for a white innovations sequence.

In the discussions and results presented above the (2,2) element of the sample correlation matrix is not referred to. This is due to the fact that its behaviour is very similar to the behaviour of the (1,1) element.

In continuation of the first-order assessment of the discrimination capability of the SSC approach, consider now the case of processing data from each of the two hypotheses using an alternative hypothesis filter, corresponding to the combined

74

process of signal + clutter + noise. Since the signal and clutter are uncorrelated in this set of examples, a sixth-order state space model is required for the combined process. As before, results are presented for two sets of conditions: (a) SNR = 3 dB and CNR = 6 dB; and (b) SNR = 3 dB and CNR = 20 dB. For each set of conditions the procedure described next was followed (all plots are for single-realization cases).

- A realization of the combined signal + clutter + noise process of duration $M = 2,200$ was generated and processed to design a sixth-order Kalman filter. The resulting filter is the filter for the alternative hypothesis (signal present).

- The alternative hypothesis filter was applied to a combined process sequence of duration $M = 2,200$, and the sample correlation matrix sequence of the filter output sequence was calculated. The real part of the (1,1) element of the resulting sample correlation matrix sequence is plotted in Figure 6-10 for CNR = 6 dB conditions, and Figure 6-11 for CNR = 20 dB conditions. As attested in both figures, the correlation sequences correspond to white innovations sequences, as expected (both figures show low-level energy content at the higher lags).

- The alternative hypothesis filter was applied to a clutter + noise process sequence (null hypothesis case) of duration $M = 2,200$, and the sample correlation matrix sequence of the filter output was calculated. In this case, however, the sequence is not a true innovations sequence because the filter is not optimal for this process. The real part of the (1,1) element of the

75

resulting sample correlation matrix sequence is plotted in Figure 6-12 for CNR = 6 dB conditions, and Figure 6-13 for CNR = 20 dB conditions. The correlation sequence in each of the figures corresponds to a colored process, and not to a white innovations sequence. This is the expected result.

Figures 6-10 through 6-13 do not include the imaginary part of the sample correlation sequence because it is similar to the imaginary part of the sample correlation sequence presented in the preceding figures. For this case also the behaviour of the (2,2) element is very similar to that of the (1,1) element of the sample correlation matrix sequence

These results indicate that the innovations-based detection methodology using the identification algorithm adopted in this program can discriminate between data corresponding to each of the two hypotheses. That is, a filter designed for the alternative hypothesis (signal + clutter + noise) generates a true innova⁺ ons sequence given a signal + clutter + noise channel process, and generates a colored output given a clutter + noise channel process. Analogously, a filter designed for the null hypothesis (clutter + noise) generates a true innovations sequence given a clutter + noise channel process, and generates a colored output given a signal + clutter + noise channel process.

first element of innovations covariance for null hypothesis data
using null hypothesis filter (order 4, CNR=20dB)

first element of innovations covariance for null hypothesis data
using null hypothesis filter (order 4, CNR=20dB)

Figure 6-7. Real and imaginary parts of the auto-correlation function of the (1,1) element of the innovations sequence vector for the case of null hypothesis data using the null hypothesis filter (CNR = 20 dB conditions).

77

first element of innovations covariance for alternative
hypothesis data using null hypothesis filter (order 4, CNR=6dB)

first element of innovations covariance for alternative hypothesis
data using null hypothesis filter (order 4, CNR=6dB)

Figure 6-8. Real and imaginary parts of the auto-correlation function of the (1,1) element of the filter output vector for the case of alternative hypothesis data using the null hypothesis filter (CNR = 6 dB conditions).

78

Figure 6-9. Real and imaginary parts of the auto-correlation function of the (1,1) element of the filter output vector for the case of alternative hypothesis data using the null hypothesis filter (CNR = 20 dB conditions).

79

first element of innovations covariance for alternative
hypothesis data using alternative hypothesis filter (order 6)

Figure 6-10. Real part of the auto-correlation function of the
(1,1) element of the innovations sequence vector for the case of
alternative hypothesis data using alternative hypothesis filter
(CNR = 6 dB conditions).

first element of innovations covariance for alternative hypothesis
data using alternative hypothesis filter (order 6, CNR=20dB)
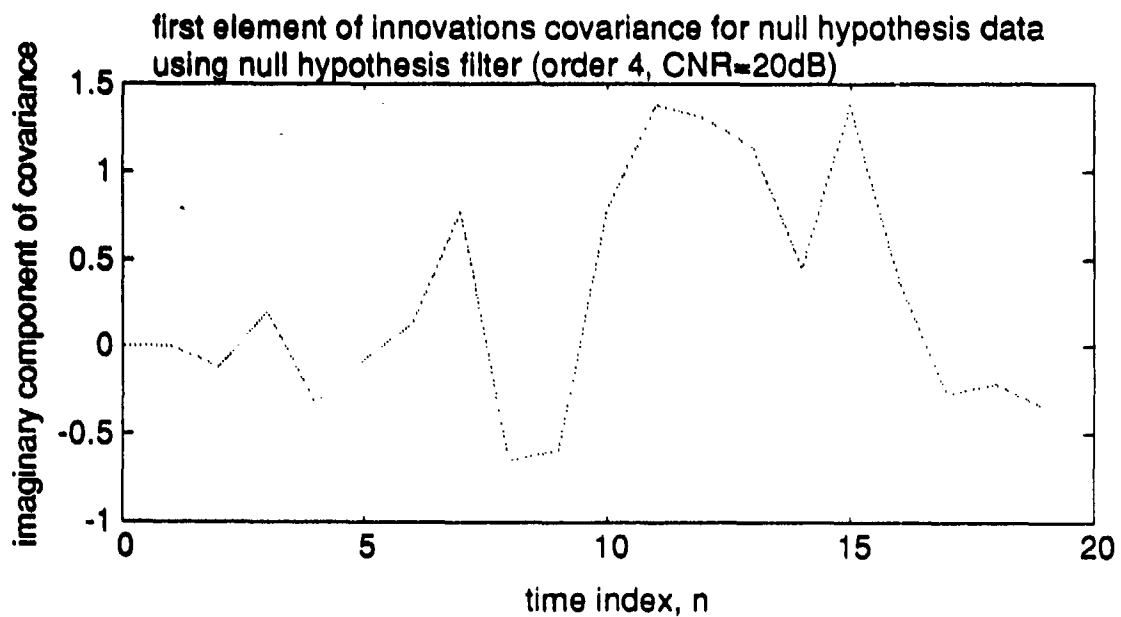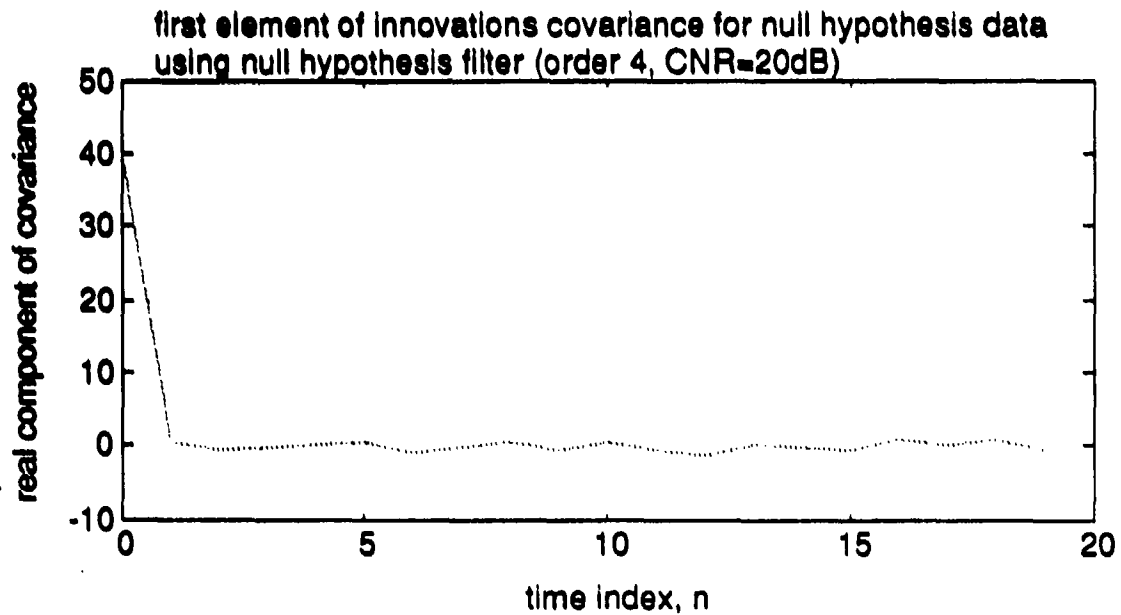
Figure 6-11. Real part of the auto-correlation function of the
(1,1) element of the innovations sequence vector for the case of
alternative hypothesis data using alternative hypothesis filter
(CNR = 20 dB conditions).

80

first element of innovations covariance for null hypothesis data using alternative hypothesis filter (order 6)
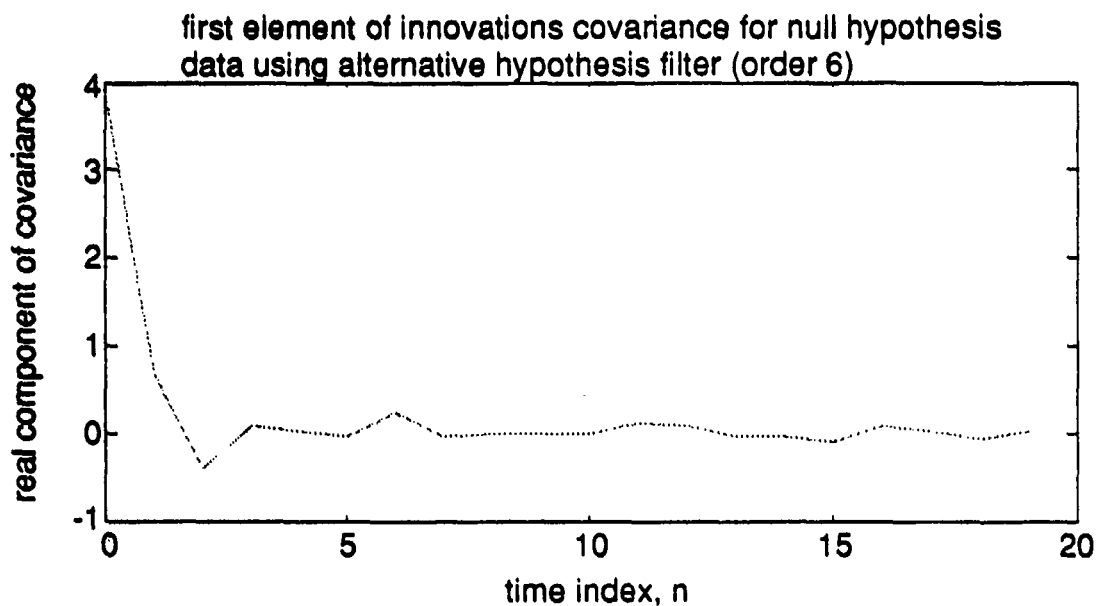
Figure 6-12. Real part of the auto-correlation function of the
(1,1) element of the filter output vector for the case of null
hypothesis data using the alternative hypothesis filter
(CNR = 6 dB conditions).

first element of innovations covariance for null hypothesis data using alternative hypothesis filter (order 6, CNR=20dB)
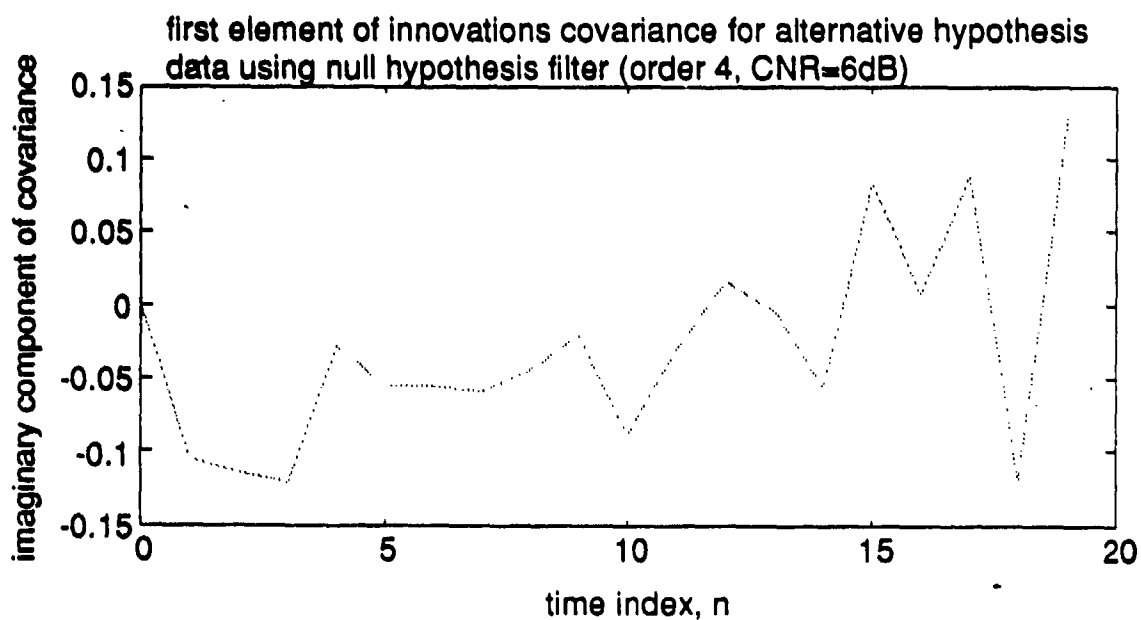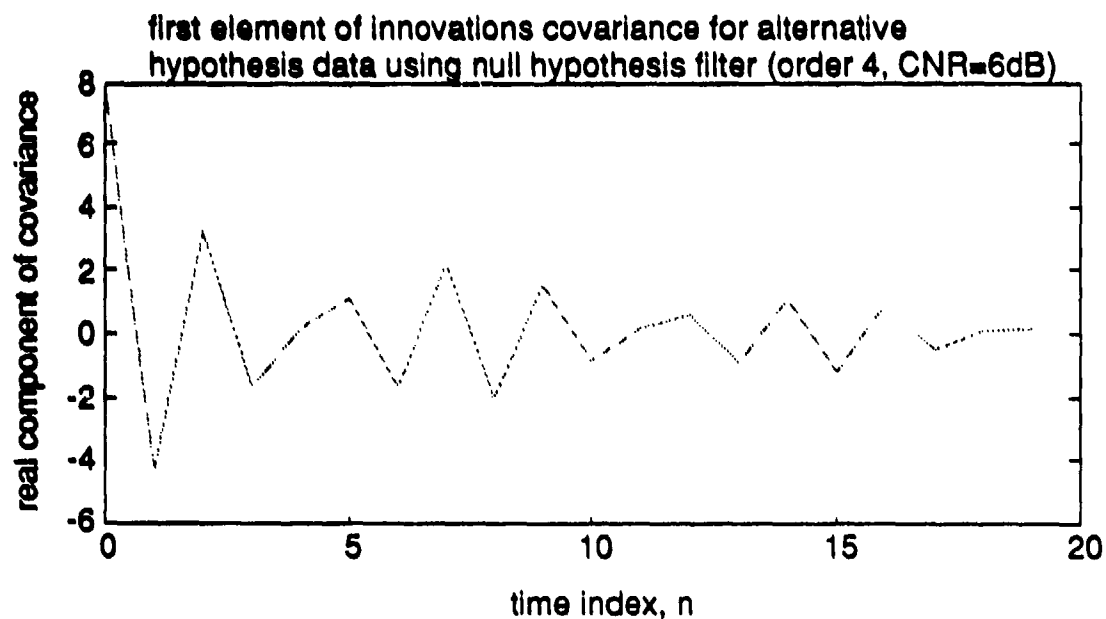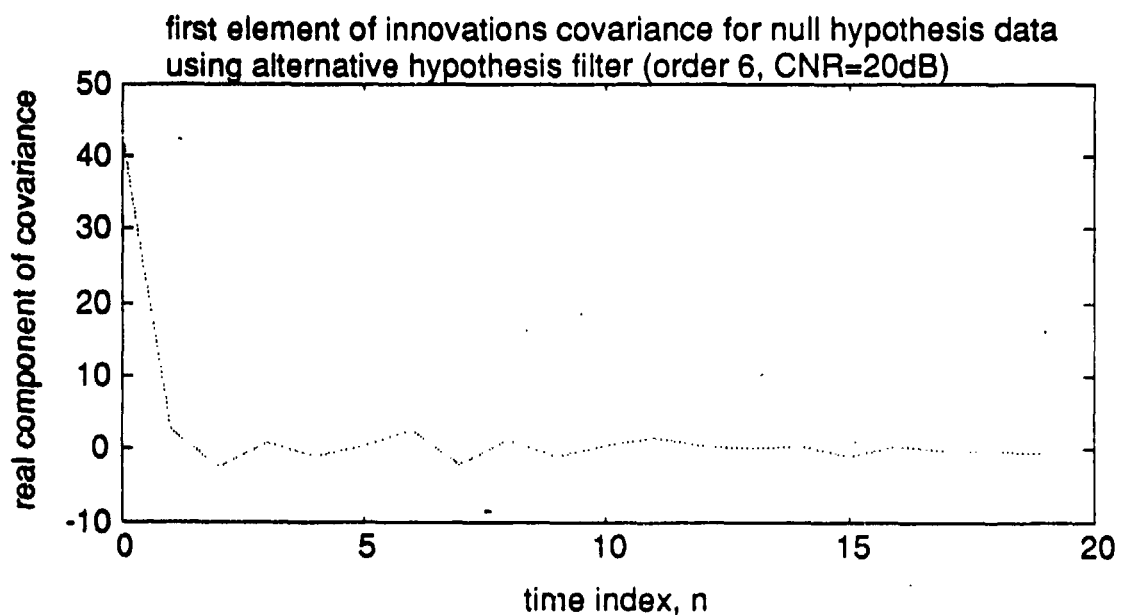
Figure 6-13. Real part of the auto-correlation function of the
(1,1) element of the filter output vector for the case of null
hypothesis data using the alternative hypothesis filter
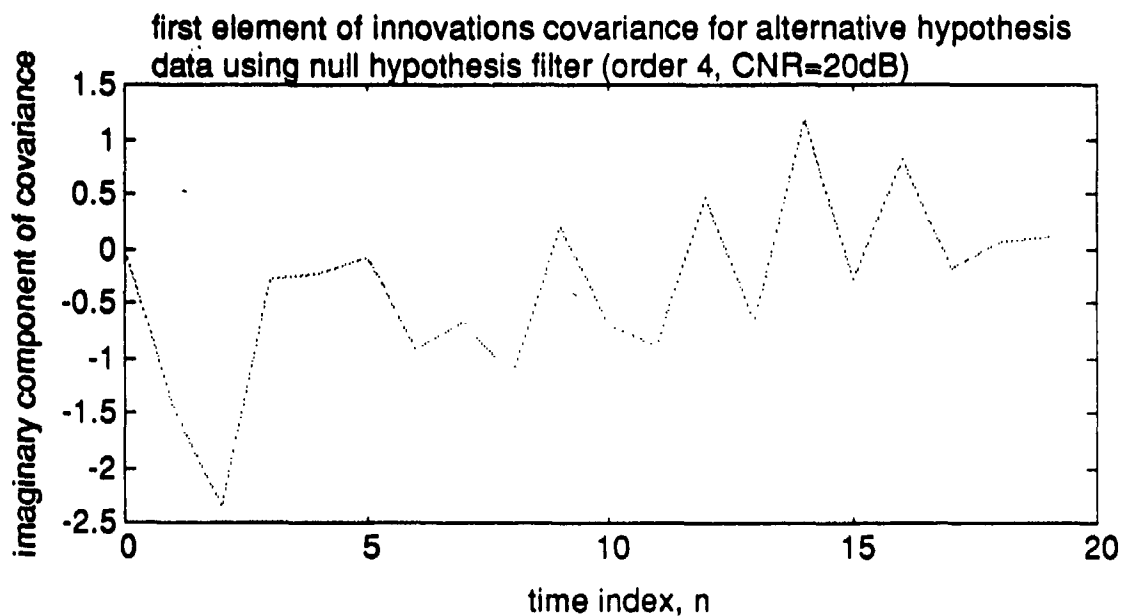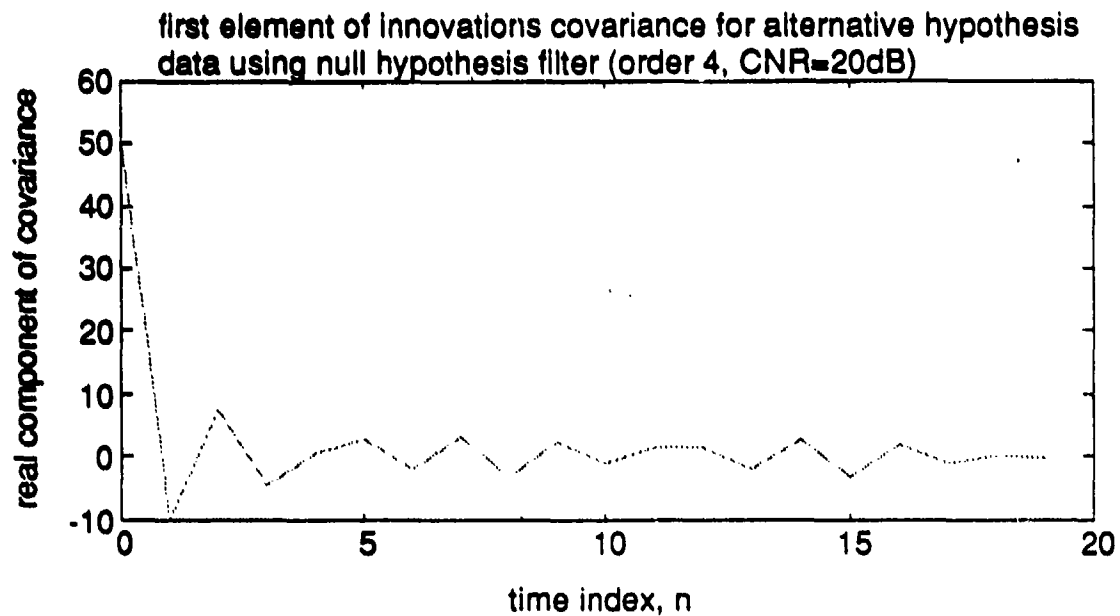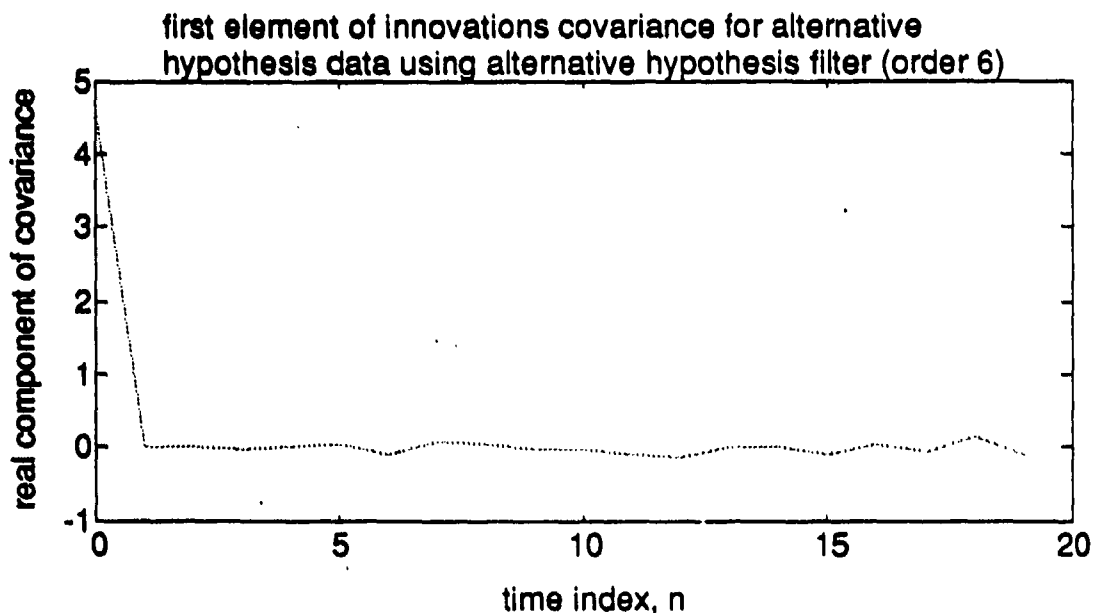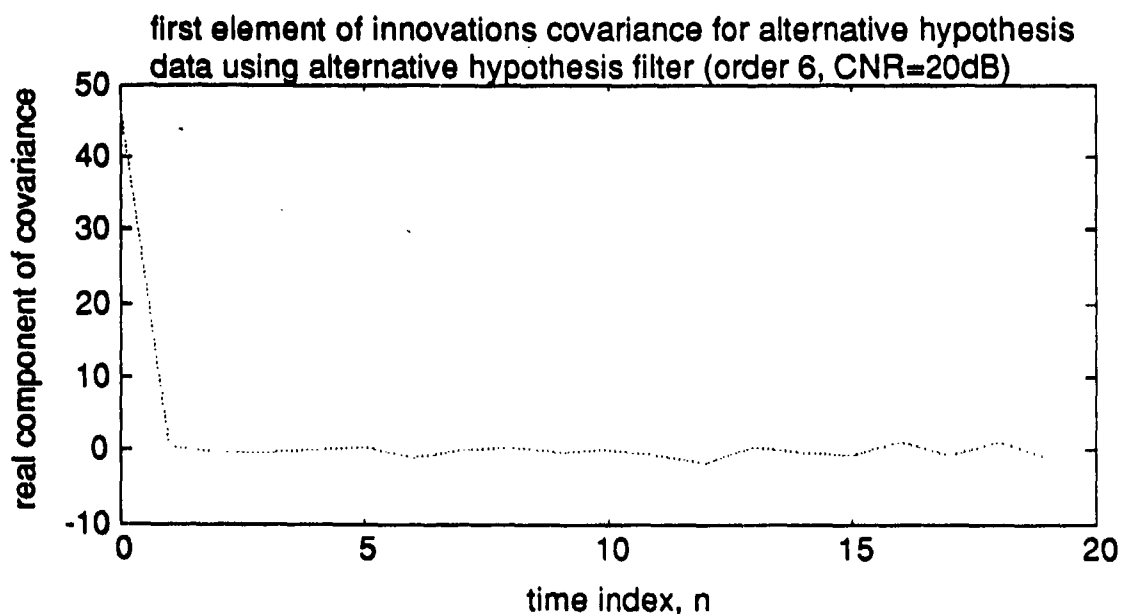(CNR = 20 dB conditions).

7.0 CONCLUSIONS AND RECOMMENDATIONS

The work carried out in this program emphasized the development and analysis of a state space methodology and algorithm for the model-based multichannel detection problem in the context of radar system applications. Application of state space techniques for multichannel detection in radar systems is one novel aspect of the work reported here. The state space model class is richer than the time series model class that is used often in radar system applications. And, as demonstrated in this work, the state space model class can be used to represent effectively multichannel radar signals.

Another novel aspect of the work is the utilization in the detection methodology of a new algorithm developed by Van Overschee and De Moor (1993). This algorithm was adopted in the program for multichannel radar output modeling and parameter identification. In the process, the algorithm was extended to the case of complex-valued radar system data, and an alternative derivation of the algorithm was developed which is simpler and easier to follow than the one presented in the forthcoming paper (Van Overschee and De Moor, 1993). The selected approach uses channel output data directly (as opposed to output correlation matrices) to estimate model parameters. This eliminates the large computational burden associated with the generation of the output correlation matrix sequence, and leads to reduced numerical precision (dynamic range) requirements. Furthermore, in a practical environment it may be possible to start processing the data as it is received. In contrast, techniques which require the computation of channel output correlation matrices have a built-in delay because the calculation of every lag requires availability of all the channel output sequence.

82

The Van Overschee-De Moor algorithm belongs to a class of techniques referred to as subspace methods. Subspace methods are based on decomposing the vector space spanned by the channel outputs into signal and noise subspaces. This decomposition is carried out with robust numerical techniques such as the SVD and the QR decomposition. Thus, the algorithm offers numerical and performance advantages over other techniques.

A computer simulation was developed to validate the algorithm and methodology, and to serve as a testbed for evaluation of the algorithm in radar system applications. The simulation can be exercised with internally-generated sample multichannel output data, or with externally-provided data. Extensive tests were carried out to validate the code.

Simulation-based analyses carried out to date demonstrate the feasibility of the SSC state space approach for multichannel identification and detection in radar system applications. The algorithm has demonstrated the capability to discriminate between signal plus clutter plus noise and clutter plus noise in an innovations-based detection algorithm formulation for the multichannel detection problem. Several cases have been analyzed at various SNR and CNR levels, and in all cases simulated thus far discrimination has been demonstrated.

In the process of completing the work reported here several areas have been identified for further research and development in a Phase II of this program. These areas are summarized below.

Processor Requirements Definition

Determination of the true potential of the SSC approach for radar system applications requires the establishment of a detailed

set of requirements for various radar problems such as space/time processing in a radar array system, the fusion of data from multiple distinct radar systems, and the fusion of multiple radar measurements (dual-polarization measurements, for example). Also, detailed sets of requirements need to be defined for other application areas such as hydrological systems, seismic event detection, and medical technology.

Additional Analyses and Detailed Algorithm Formulation

The analyses listed below are required to generate a detailed algorithm definition for the requirements, and to provide a precise assessment of the SSC approach in the context of the requirements.

- The innovations model matrix parameters F, Γ, and H can be estimated using different equations. Some of the alternative equations can exhibit bias errors, but may be simpler to implement. These alternatives need to be evaluated and traded.

- Model order selection criteria for on-line and off-line decisions need to be evaluated and traded. This includes the diagonal values $\{s_i\}$ of matrix S_L, the square of the $\{s_i\}$, and their normalized running sums.

- The steady-state Kalman filter was used in this Phase I to generate the innovations sequence. Alternatively, the time-varying Kalman filter can be used. The loss in performance, if any, incurred by using the steady-state approximation needs to be evaluated. A related issue is the duration of the transient effect in the case of the steady-state filter.

84

- Key implementation parameters need to be established. This includes the minimum required channel output sequence duration, and the row dimension of the Hankel output data matrix.

- Identification and detection performance should be compared with that of other methods. This includes state space methods that operate on output correlation matrix data, and methods based on AR models.

- A detailed algorithmic approach to the implementation of the QR decomposition and the QSVD needs to be defined. In this context, a new decomposition for rectangular matrices introduced by Stewart (1992) should be reviewed for possible utilization in the SSC approach. This decomposition is related to the QR decomposition and to the SVD, and can be updated recursively in a simple manner. The recursive feature is attractive for reducing the computational load.

Once these technical issues are addressed, a detailed architecture design can be defined.

Real-Time Processor Architecture Design

A real-time implementation architecture for the algorithm needs to be developed, and a candidate hardware implementation should be identified. Specifically, the following issues should be addressed.

- Generation of an architecture design that best meets the features of the detailed algorithm design and the

85

established processor requirements. The result may be
an architecture with features different from those in
existing processors, and which is likely to consist of
various fundamental architectures (systolic; vector;
parallel arrays; etc.).

- Analysis of state-of-the-art processors to determine
  which contemporary and next-generation VLSI components
  best match the optimized architecture design and the
  requirements.

In addressing these issues the emphasis should be on the most
computation-intensive tasks of the SSC multichannel algorithm
reported here.

Processor Development System Design

A processor development system should be designed and
installed to serve as a testbed for the development of detection
and identification methodologies and algorithms. The system
should be applicable for on-line laboratory experimentation and
for off-line processing of data collected using operational radar
systems. Availability of such a development system will speed up
significantly the algorithm development work at both SSC (during
Phase II) and RL (after delivery upon program conclusion) because
the generation of detection results require Monte Carlo analyses
and simulations.

APPENDIX A.   STATE SPACE REPRESENTATION OF TIME SERIES MODELS

Consider a discrete-time, time-invariant, complex-valued, zero-mean, random process $\{\underline{x}(n)\}$ defined as the output of the following state space system model

(A-1a) $\qquad \underline{y}(n+1) = F\underline{y}(n) + G\underline{u}(n)$

(A-1b) $\qquad \underline{x}(n) = H^H\underline{y}(n) + D^H\underline{w}(n)$

Vector recursive processes such as moving-average (MA), auto-regressive (AR), and auto-regressive moving-average (ARMA) processes can be modeled with state variable models (SVMs) of the form (A-1). The discussion herein is limited to the particular case where the matrix coefficients of the recursion are square matrices, and the number of output coefficients is equal to the number of input coefficients. The generation of a minimal-order SVM for a vector recursive process involves the properties of polynomial matrix pairs and canonical forms for multiple input, multiple output SVMs.

In contrast, minimal-order SVMs for scalar recursive processes (MA, AR, ARMA) can be generated in a straightforward manner given the recursion coefficients. The SVM generic form appropriate for modeling scalar recursive processes is

(A-2a) $\qquad \underline{y}(n+1) = F\underline{y}(n) + \underline{g}u(n)$

(A-2b) $\qquad x(n) = \underline{h}^H\underline{y}(n) + d^*w(n)$

This SVM is a single-input, single-output system.

A.1 Scalar MA Process Model

A scalar MA process of order M is defined as

$$x(n) = \sum_{k=0}^{M} b_k^* u(n-k)$$

$$x(n) = b_0^* u(n) + b_1^* u(n-1) + b_2^* u(n-2) + \ldots + b_M^* u(n-M)$$

where $\{u(n)\}$ is a zero-mean white noise sequence. This recursion can be modeled with a state-space system of the form (A-2) with input sequence $\{u(n)\}$, and state vector with elements that are determined by the input sequence,

$$\underline{y}(n) = \begin{bmatrix} y_1(n) \\ \vdots \\ y_{M-1}(n) \\ y_M(n) \end{bmatrix} = \begin{bmatrix} u(n-1) \\ \vdots \\ u(n-M+1) \\ u(n-M) \end{bmatrix} \qquad \forall\, n$$

The output noise sequence is also equal to the input noise sequence,

$$w(n) = u(n) \qquad \forall\, n$$

which means that the input and output noise sequences in the state space model are completely correlated. Model parameters (F, $\underline{g}$, $\underline{h}$, d) are defined as

$$F = \begin{bmatrix} 0 & 0 & \cdots\cdots & 0 & 0 \\ 1 & 0 & \cdots\cdots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \cdots\cdots 1 & 0 & 0 \\ 0 & 0 & \cdots\cdots 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} & \underline{0}_M^T & \\ I_{M-1} & & \underline{0}_{M-1} \end{bmatrix}$$

88

$$g = i_1 = \begin{bmatrix} 1 \\ \underline{0}_{M-1} \end{bmatrix}$$

$$\underline{h}^H = \underline{b}^H$$

$$d^* = b_0^*$$

with $\underline{b}^H$ denoting a 1xM vector defined by M of the MA recursion coefficients,

$$\underline{b}^H = \begin{bmatrix} b_1^* & b_2^* & \ldots & b_M^* \end{bmatrix}$$

The special form of matrix F is one of the possible four variations of the so-called companion matrix form. Also, the system parameters, the quadruple $(F, \underline{g}, \underline{h}, d)$, is a variation of the so-called controllable canonical form. These forms have the minimal number of non-zero elements (whereby the name "canonical") of all possible SVMs that model the scalar MA process.

Note that the definition of the state vector $\underline{y}(n)$ in terms of the sequence $\{u(n)\}$ inherently defines the initial condition vector, $\underline{y}(0)$. Once the initial condition vector is defined, the state propagation, Equation (A-2a), provides for continued generation of the output process.

Verification of the above-defined model proceeds as follows. The form of matrix F provides for continued "scrolling" of the input noise sequence as elements of $\underline{y}(n)$, for all n. Validation of the model follows from (A-2b) and the definition of $\underline{h}$, $w(n)$, and $\underline{y}(n)$. That is,

$$x(n) = \underline{h}^H \underline{y}(n) + d^* w(n) = \underline{b}^H \underline{y}(n) + b_0^* u(n)$$

Expanding the term $\underline{b}^H \underline{y}(n)$, and substitution of the definition of $\underline{y}(n)$ in terms of the sequence $\{u(n)\}$ results in

$$x(n) = b_0^* u(n) + b_1^* u(n-1) + b_2^* u(n-2) + \ldots + b_M^* u(n-M)$$

which is the MA process definition. Model validation can be carried out also using the transfer function concept, as summarized next.

Consider first the derivation of the transfer function from the MA process definition. Since the MA process is a discrete-time process, the appropriate tool for the determination of the transfer function is the z-transform. Application of the z-transform to the definition of the MA model results in the expression

$$X(z) = \sum_{k=0}^{M} b_k^* z^{-k} U(z)$$

where z denotes the transform variable, and $X(z)$ and $U(z)$ are the z-transforms of the sequences $\{x(n)\}$ and $\{u(n)\}$, respectively. The transfer function for this linear system is then defined as

$$T(z) = \frac{X(z)}{U(z)} = \sum_{k=0}^{M} b_k^* z^{-k}$$

This corresponds to the transfer function of an all-zero system, as is well known.

The transfer function for a single-input, single-output state variable model (A-2) is of the form

$$T(z) = \underline{h}^H[z\underline{I} - \underline{F}]^{-1}\underline{g} + d^*$$

The particular characteristics of matrix $\underline{F}$ and vector $\underline{g}$ lead to a very simple expression for the product $[z\underline{I} - \underline{F}]^{-1}\underline{g}$; namely,

$$[z\underline{I} - \underline{F}]^{-1}\underline{g} = \frac{1}{\gamma(z)} \underline{\theta}(z)$$

where $\gamma(z)$ is the system characteristic polynomial (the determinant of matrix $[z\underline{I} - \underline{F}]$),

$$\gamma(z) = z^M$$

and $\underline{\theta}(z)$ is vector with elements of the form $\theta_i(z) = z^{i-1}$; that is,

$$\underline{\theta}^T(z) = \left[z^{M-1} \quad \cdots \quad z^2 \quad z \quad 1 \right]$$

Substitution of these expressions and of $\underline{h}^H$ and d^* in the equation for the transfer function leads to the following result

$$T(z) = \frac{\underline{h}^H\underline{\theta}(z) + d^*\gamma(z)}{\gamma(z)} = \frac{\underline{h}^H\underline{\theta}(z) + b_0^* z^M}{z^M} = z^{-M}\left[b_0^* z^M + \underline{h}^H\underline{\theta}(z)\right]$$

$$T(z) = b_0^* + b_1^* z^{-1} + b_2^* z^{-2} + \ldots + b_M^* z^{-M} = \sum_{k=0}^{M} b_k^* z^{-k}$$

This result is identical to the transfer function expression derived from the definition of the MA process.

A.2 Scalar AR Process Model

A scalar AR process of order M is defined as

$$x(n) = -\sum_{k=1}^{M} a_k^* x(n-k) + u(n)$$

$$x(n) = -a_1^* x(n-1) - a_2^* x(n-2) - \ldots - a_M^* x(n-M) + u(n)$$

where $\{u(n)\}$ is a zero-mean white noise sequence. This recursion can be modeled with a state-space system of the form (A-2) with input sequence $\{u(n)\}$, and state vector with elements that are determined by the output sequence,

$$\underline{y}(n) = \begin{bmatrix} y_1(n) \\ \vdots \\ y_{M-1}(n) \\ y_M(n) \end{bmatrix} = \begin{bmatrix} x(n-1) \\ \vdots \\ x(n-M+1) \\ x(n-M) \end{bmatrix} \qquad \forall\, n$$

The output noise sequence is equal to the input noise sequence,

$$w(n) = u(n) \qquad\qquad\qquad \forall\, n$$

This implies complete correlation between the input and output noise sequences in the SVM (as in the case of the MA model). Model parameters $(F, \underline{g}, \underline{h}, d)$ are defined as

$$F = \begin{bmatrix} -a_1^* & -a_2^* & \cdots & \cdots & \cdots & -a_{M-1}^* & -a_M^* \\ 1 & 0 & \cdots & \cdots & \cdots & 0 & 0 \\ 0 & 1 & \ddots & & & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 1 & 0 \end{bmatrix}$$

$$F = \begin{bmatrix} -a^H \\ I_{M-1} \quad \underline{0}_{M-1} \end{bmatrix}$$

$$\underline{g} = \underline{i}_1 = \begin{bmatrix} 1 \\ \underline{0}_{M-1} \end{bmatrix}$$

$$\underline{h}^H = -a^H$$

$$d^* = 1$$

with $\underline{a}^H$ denoting a vector with elements equal to the AR recursion coefficients,

$$\underline{a}^H = \begin{bmatrix} a_1^* & a_2^* & \cdots & a_M^* \end{bmatrix}$$

The system parameters quadruple $(F, \underline{g}, \underline{h}, d)$, is in controllable canonical form, as in the MA model case.

Note that the definition of the state vector $\underline{y}(n)$ in terms of the sequence $\{x(n)\}$ inherently defines the initial condition vector, $\underline{y}(0)$. Once the initial condition vector is defined, the state propagation, Equation (A-2a), provides for continued generation of the output process.

Verification of the above-defined model proceeds as follows. From (A-2a) and the definition of F, $\underline{y}(n)$, $\underline{y}(n+1)$, and $\underline{g}$, it follows that

$$y_M(n+1) = -a_1^* y_1(n) - a_2^* y_2(n) - \ldots - a_M^* y_M(n) + u(n)$$

$$y_M(n+1) = -\underline{a}^H \underline{y}(n) + u(n)$$

Also, it follows from (A-2b) and the definition of $\underline{h}$, $w(n)$, and $\underline{y}(n)$ that

$$x(n) = \underline{h}^H \underline{y}(n) + w(n) = -\underline{a}^H \underline{y}(n) + u(n)$$

which indicates that $x(n) = y_M(n+1)$. Then, expanding the term $-\underline{a}^H \underline{y}(n)$ and substitution of the definition of $\underline{y}(n)$ in terms of the sequence $\{x(n)\}$ results in

$$x(n) = -a_1^* x(n-1) - a_2^* x(n-2) - \ldots - a_M^* x(n-M) + u(n)$$

which is the AR process definition.

The transfer function approach can be used also to validate this SVM for scalar AR processes. Application of the Z-transform to the definition of the AR model results in the expression

$$\sum_{k=0}^{M} a_k^* z^{-k} X(z) = U(z)$$

where $a_0 = 1$ is introduced for notational simplicity, and $X(z)$ and $U(z)$ are the z-transforms of the sequences $\{x(n)\}$ and $\{u(n)\}$, respectively. The transfer function for this linear system is then defined as

$$T(z) = \frac{X(z)}{U(z)} = \frac{1}{\sum_{k=0}^{M} a_k^* z^{-k}}$$

This corresponds to the transfer function of an all-pole system, as is well known.

Consider now the transfer function for the state variable model (A-2). In the present AR process case, the system characteristic polynomial is

$$\gamma(z) = z^M + a_1^* z^{M-1} + \ldots + a_{M-1}^* z + a_M^*$$

and the particular characteristics of matrix F and vector g lead the same simple expression for the product $[zI - F]^{-1} g$; namely,

$$[zI - F]^{-1} g = \frac{1}{\gamma(z)} \theta(z)$$

where $\theta(z)$ is as defined previously. Notice that the characteristic polynomial can be expressed as

$$\gamma(z) = z^M + a^H \theta(z)$$

Substitution of these expressions and of h^H and d^* in the equation for the transfer function leads to the following result

95

$$T(z) = \frac{\underline{h}^H \underline{\theta}(z) + d^* \gamma(z)}{\gamma(z)} = \frac{-\underline{a}^H \underline{\theta}(z) + \gamma(z)}{\gamma(z)} = \frac{z^M}{\gamma(z)}$$

$$T(z) = \frac{1}{z^{-M}\gamma(z)} = \frac{1}{\sum\limits_{k=0}^{M} a_k^* z^{-k}}$$

This is identical to the transfer function expression derived from the definition of the AR process.

A.3 Scalar ARMA Process Model

A scalar ARMA process of order M is defined as

$$x(n) = -\sum_{k=1}^{M} a_k^* x(n-k) + \sum_{k=0}^{M} b_k^* u(n-k)$$

$$x(n) = -a_1^* x(n-1) - \ldots - a_{M-1}^* x(n-M+1) - a_M^* x(n-M) + b_0^* u(n) + b_1^* u(n-1) +$$

$$+ b_2^* u(n-2) + \ldots + b_M^* u(n-M)$$

where $\{u(n)\}$ is a zero-mean white noise sequence. This recursion can be modeled with a state-space system of the form (2) with input sequence $\{u(n)\}$, and output noise sequence equal to the input sequence,

$$w(n) = u(n) \qquad\qquad \forall\, n$$

This implies complete correlation between the input and output noise sequences in the SVM (as in the case of the MA and the AR models). Model parameters (F, $\underline{g}$, $\underline{h}$, d) are defined as

$$F = \begin{bmatrix} -a_1^* & -a_2^* & \cdots & \cdots & \cdots & -a_{M-1}^* & -a_M^* \\ 1 & 0 & \cdots & \cdots & \cdots & 0 & 0 \\ 0 & 1 & \ddots & & & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 1 & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 1 & 0 \end{bmatrix}$$

$$F = \begin{bmatrix} -\mathbf{a}^H \\ I_{M-1} \quad \underline{0}_{M-1} \end{bmatrix}$$

$$\mathbf{g} = \mathbf{i}_1 = \begin{bmatrix} 1 \\ \underline{0}_{M-1} \end{bmatrix}$$

$$\underline{h}^H = \underline{b}^H - b_0^* \mathbf{a}^H$$

$$d^* = b_0^*$$

Here, as in the AR case, vector $\mathbf{a}^H$ has elements equal to the AR recursion coefficients,

$$\mathbf{a}^H = \begin{bmatrix} a_1^* & a_2^* & \cdots & a_M^* \end{bmatrix}$$

and vector $\underline{b}^H$ has elements defined by M of the MA recursion coefficients,

97

$$\underline{b}^H = \begin{bmatrix} b_1^* & b_2^* & \dots & b_M^* \end{bmatrix}$$

The system parameters quadruple $(F, \underline{g}, \underline{h}, d)$, is in controllable canonical form, as in the MA and AR model cases.

State vector initial conditions, $\underline{x}(0)$, for this case are related to the input and output sequences in a more complex manner, and have to be selected appropriately. Once the initial condition vector is defined, the state propagation, Equation (A-2a), provides for continued generation of the output process.

The simplest approach to validate this model is via the transfer function approach. Application of the Z-transform to the definition of the ARMA model results in the expression

$$\sum_{k=0}^{M} a_k^* z^{-k} X(z) = \sum_{k=0}^{M} b_k^* z^{-k} U(z)$$

where, as before, $X(z)$ and $U(z)$ are the z-transforms of the sequences $\{x(n)\}$ and $\{u(n)\}$, respectively, and $a_0 = 1$ is introduced for notational simplicity. The transfer function for this linear system is then defined as

$$T(z) = \frac{X(z)}{U(z)} = \frac{\displaystyle\sum_{k=0}^{M} b_k^* z^{-k}}{\displaystyle\sum_{k=0}^{M} a_k^* z^{-k}} = \frac{\displaystyle\sum_{k=0}^{M} b_k^* z^{M-k}}{\displaystyle\sum_{k=0}^{M} a_k^* z^{M-k}}$$

where the two polynomial ratio expressions (corresponding to inverse powers of Z or direct powers of Z) are equivalent, as indicated. This is a transfer function with both poles and zeros, as expected for an ARMA process.

98

Consider now the transfer function for the state variable model (A-2). For an ARMA process the system characteristic polynomial is

$$\gamma(z) = z^M + a_1^* z^{M-1} + \ldots + a_{M-1}^* z + a_M^*$$

which is equal to that for an AR process SVM model. As in the other two cases,

$$[zI - F]^{-1} g = \frac{1}{\gamma(z)} \underline{\theta}(z)$$

given the particular features of matrix F and vector g ($\underline{\theta}(z)$ is as defined previously). Notice also that, as in the AR process case, the characteristic polynomial can be expressed as

$$\gamma(z) = z^M + \underline{a}^H \underline{\theta}(z)$$

Substitution of these expressions and of $\underline{h}^H$ and d^* in the equation for the transfer function leads to the following result

$$T(z) = \frac{\underline{h}^H \underline{\theta}(z) + d^* \gamma(z)}{\gamma(z)} = \frac{\left(\underline{b}^H - b_0^* \underline{a}^H\right) \underline{\theta}(z) + b_0^* \gamma(z)}{\gamma(z)} = \frac{b_0^* z^M + \underline{b}^H \underline{\theta}(z)}{\gamma(z)}$$

It is easy to verify that this result is identical to the transfer function expression derived from the definition of the ARMA process. That is,

$$T(z) = \frac{b_0^* z^M + \underline{b}^H \underline{\theta}(z)}{\gamma(z)} = \frac{\displaystyle\sum_{k=0}^{M} b_k^* z^{M-k}}{\displaystyle\sum_{k=0}^{M} a_k^* z^{M-k}}$$

99

where $a_0 = 1$, as before.

A.4 Models for Vector Recursive Processes

Vector recursive processes of the MA, AR, and ARMA type can be represented with SVMs of the type given herein. For vector recursive processes the appropriate notation is:

MA $\quad\quad \underline{x}(n) = B_0^H \underline{u}(n) + B_1^H \underline{u}(n-1) + B_2^H \underline{u}(n-2) + \ldots + B_M^H \underline{u}(n-M)$

AR $\quad\quad \underline{x}(n) = - A_1^H \underline{x}(n-1) - A_2^H \underline{x}(n-2) - \ldots - A_M^H \underline{x}(n-M) + \underline{u}(n)$

ARMA $\quad \underline{x}(n) = - A_1^H \underline{x}(n-1) - \ldots - A_{M-1}^H \underline{x}(n-M+1) - A_M^H \underline{x}(n-M) + B_0^H \underline{u}(n) + B_1^H \underline{u}(n-1) +$

$\quad\quad\quad\quad + B_2^H \underline{u}(n-2) + \ldots + B_M^H \underline{u}(n-M)$

where each of the coefficient matrices is dimensioned JxJ. Also analogous to the scalar case, the corresponding transfer function matrices can be defined using the z-transform; which leads to

$$T_{MA}(z) = B(z)$$

$$T_{AR}(z) = A^{-1}(z)$$

$$T_{ARMA}(z) = A^{-1}(z) B(z)$$

where A(z) and B(z) are the following matrix polynomials in z,

$$A(z) = \sum_{k=0}^{M} A_k^H z^{-k}$$

$$B(z) = \sum_{k=0}^{M} B_k^H z^{-k}$$

100

with A_0 the JxJ identity matrix. The matrix pair $\{A(z), B(z)\}$ (including the cases with either $A(z) = I$ or $B(z) = I$) corresponding to a linear discrete-time system is referred to as a matrix polynomial representation or a matrix fraction description (MFD) for the system.

Departing from the time-domain definition for the vector recursive processes, the SVM for each of the three processes is of the same form as the corresponding scalar case SVM, with the following changes: a JxJ coefficient matrix in place of the corresponding coefficient scalar, a JxJ identity matrix (I_J) in place of each unit scalar, and a JxJ null matrix (O_J) in place of each zero-valued scalar. Specifically, the SVM for the ARMA vector process is:

$$
F = \begin{bmatrix}
-A_1^H & -A_2^H & \cdots & \cdots & \cdots & -A_{M-1}^H & -A_M^H \\
I_J & O_J & \cdots & \cdots & \cdots & O_J & O_J \\
O_J & I_J & \ddots & & & \vdots & \vdots \\
\vdots & \vdots & \ddots & \ddots & & \vdots & \vdots \\
\vdots & \vdots & & \ddots & \ddots & O_J & O_J \\
O_J & O_J & \cdots & \cdots & I_J & O_J & O_J \\
O_J & O_J & \cdots & \cdots & O_J & I_J & O_J
\end{bmatrix}
$$

$$G = \begin{bmatrix} I_J \\ O_J \\ \vdots \\ O_J \end{bmatrix}$$

$$H^H = \begin{bmatrix} B_1^H - A_1^H B_0^H & B_2^H - A_2^H B_0^H & \cdots & B_M^H - A_M^H B_0^H \end{bmatrix}$$

$$D^H = B_0^H$$

The SVM for the other vector processes (MA; AR) is obtained by substituting the correct values for the vector process coefficients in the above system parameters (that is, $A_i = O_J$ for an MA process; and $B_0 = I_J$ and $B_i = O_J$, $i \geq 1$ for an AR process). In all cases, the transfer function matrix is obtained from the SVM representation as

$$T(z) = H^H [zI - F]^{-1} G + D^H$$

A transfer function calculated according to this relation is equivalent to the transfer function calculated from the appropriate polynomial matrices.

The order (dimension of the state vector) of the resulting SVM for each of the three vector processes is $N = MJ$, since for each process the system matrix F consists of M block rows and M block columns, where each block in each row and column is a $J \times J$ matrix. SVM order is important for practical and computational considerations. An SVM representation is of minimal order if no other SVM representation of lower order leads to the same transfer function matrix. In terms of the system parameters (F, G, H, D), the order of the SVM representation is determined by the rank of the controllability matrix or the rank of the observability matrix, whichever is smaller. Given the form of the matrix pair

102

(F,G) for all three cases, it is easy to verify that the controllability matrix has full rank for all three cases. However, the observability matrix has a simple form only for the MA SVM. The special form of the observability matrix for the MA case considered herein (with B_M a square matrix) indicates by inspection that the rank of the observability matrix is equal to MJ if and only if matrix B_M has full rank. Such a simple result is not available for the AR and the ARMA SVMs. Determination of the conditions on the coefficients of the polynomial matrices $A(z)$ and $B(z)$ for AR and ARMA vector processes that lead to an SVM representation of minimal order is a difficult problem. This is due to the fact that both AR and the ARMA vector processes lead to a transfer function matrix with elements which are, in general, a ratio of polynomials in z.

Model order and related issues for matrix polynomial representations have been discussed by several researchers. The results summarized next are available in the text by Rosenbrock (1970). Consider the matrix polynomial representation of a system, and assume that the determinant of $A(z)$ is different from zero to eliminate pathological cases. For an AR vector process, the order of the system is given by the degree of the determinant of $A(z)$. Thus, the SVM representation presented herein for vector AR processes is of minimal order if the determinant of $A(z)$ (with $A_0 = I_J$) has degree equal to MJ.

Several definitions need to be introduced prior to stating the relevant results regarding minimal order for ARMA vector processes. A square polynomial matrix is said to be regular when the matrix coefficient of the highest power of z is non-singular. The determinant of a regular polynomial matrix has maximum possible degree. A square polynomial matrix is said to be unimodular if its determinant is a non-zero constant. Unimodular

polynomial matrices have an inverse which is also a polynomial matrix. As an example, the polynomial matrix

$$Q(z) = Q_0 + Q_1 z^{-1} = \begin{bmatrix} 1 + z^{-1} & 3 + z^{-1} \\ 2 + z^{-1} & 4 + z^{-1} \end{bmatrix}$$

is unimodular because the determinant of $Q(z)$ is equal to -2. Notice that the inverse of $Q(z)$ is also a polynomial matrix,

$$Q^{-1}(z) = \frac{1}{-2} \begin{bmatrix} 4 + z^{-1} & -(3 + z^{-1}) \\ -(2 + z^{-1}) & 1 + z^{-1} \end{bmatrix}$$

as expected. Notice also that $Q(z)$ is not a regular matrix since Q_1 is singular.

Two polynomial matrices $A(z)$ and $B(z)$ are said to have a <u>common left divisor</u> $S(z)$ if

$$A(z) = S(z)P_A(z)$$

$$B(z) = S(z)P_B(z)$$

where $S(z)$, $P_A(z)$, and $P_B(z)$ are polynomial matrices. Finally, if all the common (left) divisors of two polynomial matrices $A(z)$ and $B(z)$ are unimodular, then the two matrices are said to be <u>relatively (left) prime</u>. That is, if $A(z)$ and $B(z)$ are relatively (left) prime, then the determinant of the polynomial matrix $S(z)$ in the above factorizations is a constant. This implies that the degree of the determinant of $P_A(z)$ is equal to the degree of the determinant of $A(z)$, and the degree of the determinant of $P_B(z)$ is equal to the degree of the determinant of $B(z)$. Furthermore, the determinant of $P_A(z)$ has no polynomial factors in common with the determinant of $P_B(z)$. A matrix polynomial pair $(A(z), B(z))$ with $A(z)$

and $B(z)$ relatively (left) prime is an _irreducible_ matrix polynomial representation for the system.

The relevant results for ARMA vector processes can be stated now. As in the AR case, for an ARMA vector process the determinant of $A(z)$ (with $A_0 = I_J$) must have degree equal to MJ for the SVM representation presented herein to be of minimal order. However, two additional conditions must be satisfied. Namely, matrix B_M must have full rank, and the polynomial matrices $A(z)$ and $B(z)$ must be relatively (left) prime. Full rank for matrix B_M implies that $B(z)$ is a regular polynomial matrix. If $A(z)$ and $B(z)$ are not relatively prime, then the order of the system is reduced by the degree of the determinant of the greatest common (left) divisor of $A(z)$ and $B(z)$. This is related to the so-called pole/zero cancelations.

APPENDIX B. QUOTIENT SINGULAR VALUE DECOMPOSITION

The quotient singular value decomposition (QSVD) is a generalization of the SVD for a matrix to the case of two general matrices with the same number of columns. As such, it is also referred to as the generalized SVD. This concept was developed by Van Loan (1976), who called it the BSVD. Paige and Saunders (1981) modified the concept and extended its applicability to general matrices. Their concept is summarized herein in the context of the multichannel detection application. From a computational viewpoint, the approach suggested by Van Overschee and De Moor (1993) is adopted. Two distinct cases are considered herein, corresponding to the two conditions that arise in the implementation of the identification algorithm (Section 3.1).

B.1 QSVD for the Matrices of Equations (3-22) and (3-23)

Consider the $J(L-1) \times J(L+1)$ matrix R_D, and the $J(L-1) \times J(L-1)$ matrix R_E defined in Equations (3-22) and (3-23), respectively. It is desired to determine the QSVD of the matrix pair consisting of the conjugate transpose of these two matrices. The procedure is described below.

The first step is to define a $2JL \times J(L-1)$ matrix A as the following concatenation of the conjugate transposes of matrices R_D and R_E:

$$(B-1) \qquad A = \begin{bmatrix} R_D^H \\ \cdots \cdots \\ R_E^H \end{bmatrix}$$

Now carry out an SVD on matrix A to get (recall that A has more rows than columns),

106

$$(B-2a) \qquad A = U_A \begin{bmatrix} S_A \\ [0] \end{bmatrix} V_A^H = \begin{bmatrix} U_{1A} & U_{3A} & U_{5A} \\ U_{2A} & U_{4A} & U_{6A} \end{bmatrix} \begin{bmatrix} S_{1A} & [0] \\ [0] & [0] \\ [0] & [0] \end{bmatrix} V_A^H$$

$$(B-2b) \qquad S_A = \begin{bmatrix} S_{1A} & [0] \\ [0] & [0] \end{bmatrix}$$

In this decomposition the unitary matrix U_A is $J(L+1) \times J(L+1)$, matrix S_A is $J(L-1) \times J(L-1)$, and the unitary matrix V_A is $J(L-1) \times J(L-1)$. Matrix S_A is diagonal, with real-valued non-negative elements along the diagonal arranged in decreasing order of magnitude (the largest-valued element occupies the $(1,1)$ position). The diagonal elements of matrix S_A are the singular values of matrix A. The rank of matrix A, denoted herein as K_A = rank(A), is equal to the number of non-zero singular values. These non-zero singular values are the diagonal elements of the $K_A \times K_A$ matrix S_{1A}. If matrix A is full-rank, then S_{1A} becomes S_A. In most cases involving random processes the singular values of A will be non-zero, although there may be a large dynamic range between the largest and the smallest singular values. Alternatively, the singular values may appear in groups, with a significant variation in dynamic range between the groups of singular values. Such situations are indicative of an effective rank $K_A < J(L-1)$.

The row partition of matrix U_A into submatrices is the same as the row partition in the concatenation (B-1), and the column partitioning of U_A corresponds to the row partitioning of matrix S_A. That is, matrix U_{1A} is $J(L+1) \times K_A$ and matrix U_{2A} is $J(L-1) \times K_A$. With this partitioning, Equation (B-2a) is equivalent to the following expression,

107

$$\text{(B-3)} \qquad A = \begin{bmatrix} U_{1A} & [0] \\ U_{2A} & [0] \end{bmatrix} \begin{bmatrix} S_{1A} & [0] \\ [0] & [0] \end{bmatrix} V_A^H = \begin{bmatrix} U_{1A} & [0] \\ U_{2A} & [0] \end{bmatrix} \begin{bmatrix} S_{1A} & [0] \\ [0] & I_M \end{bmatrix} V_A^H$$

where $M = J(L-1)-K_A$. This expression is obtained by first cutting off the last block column of matrix U_A and the corresponding block row of matrix S_A. Next, U_{3A} and U_{4A} are replaced with zeros, which allows placing an identity matrix of dimensions $M = J(L-1)-K_A$ into the lower right-hand-corner of matrix S_A. None of these modifications alters the numerical value of the expression.

The next step in the computation of the QSVD is to carry out an SVD on each of the matrices U_{1A} and U_{2A}. The resulting SVDs can be expressed as

$$\text{(B-4)} \qquad U_{1A} = U_{L-1} \, S_{L-1} \, V_{U1A}^H$$

$$\text{(B-5)} \qquad U_{2A} = V_{L-1} \, T_{L-1} \, V_{U2A}^H$$

In the first decomposition, U_{L-1} is a $J(L+1) \times J(L+1)$ unitary matrix, S_{L-1} is $J(L+1) \times K_A$, and V_{U1A} is a $K_A \times K_A$ unitary matrix. Matrix S_{L-1} is zero except for real-valued, non-negative elements along the main diagonal. The non-negative elements of S_{L-1} are arranged in decreasing order of magnitude, with the largest-valued element occupying the $(1,1)$ position and having value less than or equal to unity. In the second decomposition, V_{L-1} is a $J(L-1) \times J(L-1)$ unitary matrix, T_{L-1} is $J(L-1) \times K_A$, and V_{U2A} is a $K_A \times K_A$ unitary matrix. Matrix T_{L-1} is zero except for real-valued, non-negative elements along the main diagonal. The non-negative elements of T_{L-1} are arranged in increasing order of magnitude, with the smallest-valued element occupying the $(1,1)$ position and having value greater than or equal to zero. The largest-valued element of T_{L-1} is the (K_A, K_A)th element, and its value is less than or equal to unity. Notice that the arrangement of the elements of T_{L-1} along

108

the main diagonal is the reverse of the conventional SVD. This deviation, however, is essential to the Paige and Saunders (1981) definition of the QSVD. Let $\{s_i \mid i = 1, 2, \ldots, K_A\}$ denote the main diagonal elements of S_{L-1}, and let $\{t_i \mid i = 1, 2, \ldots, K_A)\}$ denote the main diagonal elements of T_{L-1}. The above-stated conditions on these elements are summarized as:

(B-6) $1 \geq s_1 \geq s_2 \geq \ldots \geq s_{K_A} \geq 0$

(B-7) $0 \leq t_1 \leq t_2 \leq \ldots \leq t_{K_A} \leq 1$

Paige and Saunders (1981) have shown also that these elements satisfy the following constraint,

(B-8) $s_i^2 + t_i^2 = 1$ $\qquad\qquad\qquad\qquad\qquad\qquad$ $i = 1, 2, \ldots, K_A$

This constraint is valid only if the singular values satisfy (B-6) and (B-7). The pairs of values (s_i, t_i) are called the __singular value pairs__ of matrices R_D and R_E.

Based on Equations (B-6) and (B-7) and on the orthogonality property of matrix U_A it is possible to show that

(B-9) $V_{U1A} = V_{U2A} = V_{UA}$

Then, substituting this result into Equations (B-4) and (B-5), and in turn substituting these into Equation (B-3) leads to

(B-10) $A = \begin{bmatrix} U_{L-1}S_{L-1} & [0] \\ V_{L-1}T_{L-1} & [0] \end{bmatrix} \begin{bmatrix} V_{UA} & [0] \\ [0] & I_M \end{bmatrix} \begin{bmatrix} S_{1A} & [0] \\ [0] & I_M \end{bmatrix} V_A^H$

Now define a $J(L-1) \times J(L-1)$ matrix Y_{L-1} as

109

$$(B-11) \qquad Y_{L\text{-}1}^H = \begin{bmatrix} V_{UA} & [0] \\ [0] & I_M \end{bmatrix} \begin{bmatrix} S_{1A} & [0] \\ [0] & I_M \end{bmatrix} V_A^H$$

Substitute for $Y_{L\text{-}1}$ in Equation (B-10) and re-arrange the submatrices of the first matrix to obtain

$$(B-12) \qquad A = \begin{bmatrix} U_{L\text{-}1} \begin{bmatrix} S_{L\text{-}1} & O_{J(L+1),M} \end{bmatrix} \\ V_{L\text{-}1} \begin{bmatrix} T_{L\text{-}1} & O_{J(L-1),M} \end{bmatrix} \end{bmatrix} Y_{L\text{-}1}^H$$

The desired QSVD for the matrix pair R_D and R_E follows directly by a comparison of Equations (B-1) and (B-12),

$$(B-13) \qquad R_D^H = U_{L\text{-}1} \begin{bmatrix} S_{L\text{-}1} & O_{J(L+1),M} \end{bmatrix} Y_{L\text{-}1}^H$$

$$(B-14) \qquad R_E^H = V_{L\text{-}1} \begin{bmatrix} T_{L\text{-}1} & O_{J(L-1),M} \end{bmatrix} Y_{L\text{-}1}^H$$

If matrix A has full rank these expressions simplify to

$$(B-15) \qquad R_D^H = U_{L\text{-}1} S_{L\text{-}1} Y_{L\text{-}1}^H$$

$$(B-16) \qquad R_E^H = V_{L\text{-}1} T_{L\text{-}1} Y_{L\text{-}1}^H$$

As mentioned earlier, in most cases involving random processes matrix A will be full rank. Even if such is not the case, it appears to be better to over-estimate the rank of A rather than to under-estimate it. In fact, if the rank of matrix A is under-estimated, then the true structure of matrices $S_{L\text{-}1}$ and $T_{L\text{-}1}$ is distorted. With over-estimation of the rank of A it is still possible to determine the true model order accurately, but under-estimation of the rank of A effectively places an upper bound on the attainable model order and this bound could be less than the true model order (see Section 3.2).

110

Consider matrix S_{L-1} for the general case, where $rank(A) = K_A \leq J(L-1)$. The structure of this matrix is determined by the true order of a state space representation of the process being modeled. Specifically, for the case where the model order is $N < K_A$, matrix S_{L-1} is of the form

$$\text{(B-17)} \qquad S_{L-1} = \begin{bmatrix} S_{L-1}^{(1)} & [0] \\ [0] & S_{L-1}^{(2)} \\ [0] & [0] \end{bmatrix} = \begin{bmatrix} S_{L-1}^{(1)} & [0] \\ [0] & [0] \\ [0] & [0] \end{bmatrix}$$

Here $S_{L-1}^{(1)}$ is an $N \times N$ diagonal matrix, and $S_{L-1}^{(2)}$ is a $(K_A-N) \times (K_A-N)$ matrix with possible non-zero elements only along the main diagonal. As inferred by Equation (B-17), $S_{L-1}^{(2)}$ is a null matrix when the model order $N < K_A$. In practical situations where randomness is present, the diagonal elements of matrix $S_{L-1}^{(2)}$ are not equal to zero, but they are significantly smaller than the diagonal elements of $S_{L-1}^{(1)}$.

B.2 QSVD for the Matrices of Equations (3-20) and (3-21)

Consider the JLxJL matrix R_B, and the JLxJL matrix R_C defined in Equations (3-20) and (3-21), respectively. It is desired to determine the QSVD of the matrix pair consisting of the conjugate transpose of these two matrices. Since the approach is analogous to the preceding section, only the key steps and definitions are given below.

As before, define a 2JLxJL matrix B as the following concatenation of the conjugate transposes of matrices R_B and R_C:

$$\text{(B-18)} \qquad B = \begin{bmatrix} R_B^H \\ \cdots\cdots \\ R_C^H \end{bmatrix}$$

Now carry out an SVD on matrix B to get (recall that B has more rows than columns),

$$\text{(B-19a)} \qquad B = U_B \begin{bmatrix} S_B \\ [0] \end{bmatrix} V_B^H = \begin{bmatrix} U_{1B} & U_{3B} & U_{5B} \\ U_{2B} & U_{4B} & U_{6B} \end{bmatrix} \begin{bmatrix} S_{1B} & [0] \\ [0] & [0] \\ [0] & [0] \end{bmatrix} V_B^H$$

$$\text{(B-19b)} \qquad S_B = \begin{bmatrix} S_{1B} & [0] \\ [0] & [0] \end{bmatrix}$$

In this decomposition the unitary matrix U_B is $2JL \times 2JL$, matrix S_B is $2JL \times JL$, and the unitary matrix V_B is $JL \times JL$. Matrix S_B is diagonal, with real-valued non-negative elements along the diagonal arranged in decreasing order of magnitude (the largest-valued element occupies the (1,1) position). The diagonal elements of matrix S_B are the singular values of matrix B, and the rank of matrix B, denoted herein as $K_B = \text{rank}(B)$, is equal to the number of non-zero singular values. These non-zero singular values are the diagonal elements of the $K_B \times K_B$ matrix S_{1B}. If matrix B is full-rank, then S_{1B} becomes S_B. Analogous to the prior case, for random processes the effective rank of matrix B is $K_B \leq JL$.

As before, Equation (B-19a) can be converted to the following equivalent form,

$$\text{(B-20)} \qquad B = \begin{bmatrix} U_{1B} & [0] \\ U_{2B} & [0] \end{bmatrix} \begin{bmatrix} S_{1B} & [0] \\ [0] & [0] \end{bmatrix} V_B^H = \begin{bmatrix} U_{1B} & [0] \\ U_{2B} & [0] \end{bmatrix} \begin{bmatrix} S_{1B} & [0] \\ [0] & I_M \end{bmatrix} V_B^H$$

$$\text{(B-20)} \qquad B = \begin{bmatrix} U_{1B} & [0] \\ U_{2B} & [0] \end{bmatrix} \begin{bmatrix} S_{1B} & [0] \\ [0] & [0] \end{bmatrix} V_B^H = \begin{bmatrix} U_{1B} & [0] \\ U_{2B} & [0] \end{bmatrix} \begin{bmatrix} S_{1B} & [0] \\ [0] & I_M \end{bmatrix} V_B^H$$

where $M = JL-K_B$. The SVD of the $JLxK_B$ matrix U_{1B} and the SVD of the $JLxK_B$ matrix U_{2B} are expressed as

$$\text{(B-21)} \qquad U_{1B} = U_L S_L V_{U1B}^H$$

$$\text{(B-22)} \qquad U_{2B} = V_L T_L V_{U2B}^H$$

In the first decomposition, U_L is a $JLxJL$ unitary matrix, S_L is $JLxK_B$, and V_{U1B} is a K_BxK_B unitary matrix. Matrix S_L is zero except for real-valued, non-negative elements along the main diagonal. The non-negative elements of S_L are arranged in decreasing order of magnitude, with the largest-valued element occupying the (1,1) position and having value less than or equal to unity. In the second decomposition, V_L is a $JLxJL$ unitary matrix, T_L is $JLxK_B$, and V_{U2B} is a K_BxK_B unitary matrix. Matrix T_L is zero except for real-valued, non-negative elements along the main diagonal. The non-negative elements of T_L are arranged in increasing order of magnitude, with the smallest-valued element occupying the (1,1) position and having value greater than or equal to zero. The largest-valued element of T_L is the (K_B, K_B)th element, and its value is less than or equal to unity. As in the prior case, the arrangement of the elements of T_L along the main diagonal is essential to this definition of the QSVD. The non-zero elements of S_L and T_L are the singular value pairs of matrices R_B and R_C, and they satisfy conditions identical to (B-6) through (B-8) with K_A replaced by K_B.

Given the conditions satisfied by the singular value pairs and given the orthogonality property of matrix U_B it is possible to show that

113

(B-23) $V_{U1B} = V_{U2B} = V_{UB}$

Then, substituting this result into Equations (B-21) and (B-22), and in turn substituting these into Equation (B-20) leads to

$$(B-24) \qquad B = \begin{bmatrix} U_L S_L & [0] \\ V_L T_L & [0] \end{bmatrix} \begin{bmatrix} V_{UB} & [0] \\ [0] & I_M \end{bmatrix} \begin{bmatrix} S_{1B} & [0] \\ [0] & I_M \end{bmatrix} V_B^H$$

Now define a JLxJL matrix Y_L as

$$(B-25) \qquad Y_L^H = \begin{bmatrix} V_{UB} & [0] \\ [0] & I_M \end{bmatrix} \begin{bmatrix} S_{1B} & [0] \\ [0] & I_M \end{bmatrix} V_B^H$$

Substitute for Y_L in Equation (B-24) and re-arrange the submatrices of the first matrix to obtain

$$(B-26) \qquad B = \begin{bmatrix} U_L [S_L & O_{JL,M}] \\ V_L [T_L & O_{JL,M}] \end{bmatrix} Y_L^H$$

The desired QSVD for the matrix pair R_B and R_C follows directly by a comparison of Equations (B-18) and (B-26),

$$(B-27) \qquad R_B^H = U_L [S_L \quad O_{JL,M}] Y_L^H$$

$$(B-28) \qquad R_C^H = V_L [T_L \quad O_{JL,M}] Y_L^H$$

If matrix B has full rank, as can be expected in most cases where the processes are random, these expressions simplify to

$$(B-29) \qquad R_B^H = U_L S_L Y_L^H$$

114

The comments made earlier regarding the rank of matrix **A** apply equally to the rank of matrix **B**. In particular, it is better to over-estimate the rank of **B** rather than to under-estimate it.

Consider matrix S_L for the cases where rank(B) = $K_B \le JL$. The structure of this matrix is determined by the true order of a state space representation of the process being modeled. In fact, model order can be determined by examining the diagonal elements of S_L (see Section 3.2). Model order can be determined also from the diagonal elements of matrix S_{L-1}. However, it is preferable to use matrix S_L for model order determination because this matrix is generated by the QSVD of two matrices with JL rows. Such a QSVD is more robust numerically than the QSVD for S_{L-1}, which is a QSVD for two matrices with J(L-1) rows. For the cases where the model order is $N < K_B$, matrix S_L is of the form

$$\text{(B-31)} \qquad S_L = \begin{bmatrix} S_L^{(1)} & [0] \\ [0] & S_L^{(2)} \\ [0] & [0] \end{bmatrix} = \begin{bmatrix} S_L^{(1)} & [0] \\ [0] & [0] \\ [0] & [0] \end{bmatrix}$$

Here $S_L^{(1)}$ is an NxN diagonal matrix, and $S_L^{(2)}$ is a $(K_B-N) \times (K_B-N)$ matrix with possible non-zero elements only along the main diagonal. As inferred by Equation (B-31), $S_L^{(2)}$ is a null matrix when the model order $N < K_B$. When random processes are being modeled the diagonal elements of matrix $S_L^{(2)}$ are not equal to zero, but their magnitude is smaller than the diagonal elements of $S_L^{(1)}$. In such cases the relative numerical value of the elements along the main diagonal of S_L determines the cut-off point, and consequently, the model order (Section 3.2).

115

REFERENCES

H. Akaike

(1974) "Stochastic Theory of Minimal Realization," _IEEE Transactions on Automatic Control_, Vol. AC-19, No. 6 (December), pp. 667-674.

(1975) "Markovian Representation of Stochastic Processes by Canonical Variables," _SIAM Journal on Control_, Vol. 13, No. 1 (January), pp. 162-173.

B. D. O. Anderson and J. B. Moore

(1979) _Optimal Filtering_, Prentice-Hall, Englewood Cliffs, NJ.

K. S. Arun and S. Y. Kung

(1990) "Balanced Approximation of Stochastic Systems," _SIAM Journal on Matrix Analysis and Applications_, Vol. 11, No. 1 (January), pp. 42-68.

J. V. Candy

(1976) _Realization of Invariant System Descriptions From Markov Sequences_, Ph. D. Dissertation, Department of Electrical Engineering, University of Florida, Gainesville, FL.

J. J. Dongarra, C. B. Moler, J. R. Bunch, and G. W. Stewart

(1979) _LINPACK Users' Guide_, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

P. L. Faurre

(1976) "Stochastic realization algorithms," in _System Identification: Advances and Case Studies_, R. K. Mehra and D. G. Laniotis (eds.), Academic Press, New York, NY.

P. A. S. Metford and S. Haykin

(1985) "Experimental analysis of an innovations-based detection algorithm for surveillance radar," *IEE Proceedings*, Vol. 132, Pt. F, No. 1 (February), pp. 18-26.

J. H. Michels

(1991) *Multichannel Detection Using the Discrete-Time Model-Based Innovations Approach*, RL Technical Report No. RL-TR-91-269, Rome Laboratory, Griffiss AFB, NY.

(1992a) "Detection of Partially Correlated Signals in Clutter Using a Multichannel Model-Based Approach," presented at the 1992 National Telesystems Conference, May 19-20, The George Washington University, Ashburn, VA.

(1992b) *Considerations of the Error Variances of Time-Averaged Estimators for Correlated Processes*, RL Technical Report No. RL-TR-92-339, Rome Laboratory, Griffiss AFB, NY.

B. C. Moore

(1981) "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Transactions on Automatic Control*, Vol. AC-26, No. 1 (February), pp. 17-31.

A. H. Nuttall

(1976) "Multivariate Linear Predictive Spectral Analysis Employing Weighted Forward and Backward Averaging: A Generalization of Burg's Algorithm," Naval Underwater Systems Center Tech. Report No. TR-5501, New London, CT.

C. C. Paige and M. A. Saunders

(1981) "Towards a Generalized Singular Value Decomposition," *SIAM Journal on Numerical Analysis*, Vol. 18, No. 3 (June), pp. 398-405.

117

M. C. Pease
(1965) Methods of Matrix Algebra, Academic Press, New York, NY.

W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling
(1989) Numerical Recipes - The Art of Scientific Computing (FORTRAN Version), Cambridge Univ. Press, New York, NY.

M. Rangaswamy, D. D. Weiner, and J. H. Michels
(1993) "Multichannel Detection for Correlated Non-Gaussian Random Processes Based on Innovations," presented at the SPIE International Symposium on Optical Engineering and Photonics in Aerospace and Remote Sensing (Conference 1955), April 12-16, Orlando, FL.

H. H. Rosenbrock
(1970) State-space and Multivariable Theory, J. Wiley & Sons, Inc., New York, NY.

R. O. Schmidt
(1979) "Multiple emitter location and signal parameter estimation," Proceedings of the RADC Spectrum Estimation Workshop, Griffiss AFB, Rome, NY, pp. 243-258; also in IEEE Transactions on Antennas and Propagation, Vol. AP-34, No. 3 (March 1986), pp. 276-280.
(1981) A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation, Ph. D. Dissertation, Department of Electrical Engineering, Stanford Univ., Stanford, CA.

G. W. Stewart
(1992) "An Update Algorithm for Subspace Tracking," IEEE Transactions on Signal Processing, Vol. 40, No. 6 (June), pp. 1535-1541.

O. N. Strand

(1977) "Multichannel Complex Maximum Entropy (Auto-Regressive) Spectral Analysis," _IEEE Transactions on Automatic Control_, Vol. AC-22, No. 4 (August), pp. 634-640.

C. F. Van Loan

(1976) "Generalizing the Singular Value Decomposition," _SIAM Journal on Numerical Analysis_, Vol. 13, No. 1 (March), pp. 76-83.

P. Van Overschee and B. De Moor

(1993) "Subspace Algorithms for the Stochastic Identification Problem," to appear in _Automatica_.

R. A. Wiggins and E. A. Robinson

(1965) "Recursive Solution to the Multichannel Filtering Problem," _Journal of Geophysical Research_, Vol. 70, No. 8 (April 15), pp. 435-441.

H. P. Zeiger and A. J. McEwen

(1974) "Approximate Linear Realizations of Given Dimension via Ho's Algorithm," _IEEE Transactions on Automatic Control_, Vol. AC-19, No. 2 (April), pg. 153.

MISSION

OF

ROME LABORATORY

Rome Laboratory plans and executes an interdisciplinary program in research, development, test, and technology transition in support of Air Force Command, Control, Communications and Intelligence (C^3I) activities for all Air Force platforms. It also executes selected acquisition programs in several areas of expertise. Technical and engineering support within areas of competence is provided to ESD Program Offices (POs) and other ESD elements to perform effective acquisition of C^3I systems. In addition, Rome Laboratory's technology supports other AFSC Product Divisions, the Air Force user community, and other DOD and non-DOD agencies. Rome Laboratory maintains technical competence and research programs in areas including, but not limited to, communications, command and control, battle management, intelligence information processing, computational sciences and software producibility, wide area surveillance/sensors, signal processing, solid state sciences, photonics, electromagnetic technology, superconductivity, and electronic reliability/maintainability and testability.

DEPARTMENT OF THE AIR FORCE
AIR FORCE RESEARCH LABORATORY (AFMC)

15 Jun 04

MEMORANDUM FOR DTIC-OCQ
                ATTN: Larry Downing
                Ft. Belvoir, VA 22060-6218

FROM: AFRL/IFOIP

SUBJECT: Distribution Statement Change

    1. The following documents (previously limited by SBIR data rights) have been reviewed and have been approved for Public Release; Distribution Unlimited:

ADB226867, "Multichannel System Identification and Detection Using Output Data Techniques", RL-TR-97-5, Vol 1.
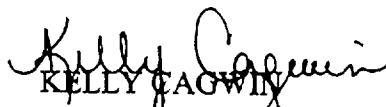
ADB176689, "Multichannel System Identification and Detection Using Output Data Techniques", RL-TR-93-141.

ADB198116, "Multichannel Detection Using Higher Order Statistics", RL-TR-95-11.

ADB232680, "Two-Dimensional Processing for Radar Systems", RL-TR-97-127.

ADB276328, "Two-Dimensional Processing for Radar Systems", AFRL-SN-RS-TR-2001-244.

    2. Please contact the undersigned should you have any questions regarding this memorandum. Thank you very much for your time and attention to this matter.

KELLY CAGWIN
STINFO Officer
Information Directorate
315-330-7094/DSN 587-7094